Faces in the mind

Featural and configural face representations in perception and imagery

Thesis

presented to the Faculty of Arts

of

the University of Zurich

for the degree of Doctor of Philosophy

by

Janek Lobmaier

of Zurich

Accepted in the winter semester 2006/2007 on the recommendation of Prof. Dr. Fred Mast and Prof. Dr. Bernhard Hess

Studentendruckerei

2007

1. FACE PROCESSING: A TOPIC OF WIDE INTEREST 2
1.1. INTRODUCTION 2
1.2. HOLISTIC HYPOTHESIS OF FACE PERCEPTION 4
1.3. SCHEMA HYPOTHESIS OF FACE PROCESSING 5
1.4. FEATURAL-CONFIGURAL HYPOTHESIS OF FACE PERCEPTION 5
1.5. THE FACE INVERSION EFFECT 7
1.6. MENTAL IMAGERY OF FACES 9
1.7. EVIDENCE FROM NEURO-IMAGING STUDIES AND NEUROPHYSIOLOGY 10
1.8. CONCLUSION 12
2. PERCEPTION OF NOVEL FACES: THE PARTS HAVE IT! 16
2.1. INTRODUCTION 16
2.2. EXPERIMENT 1 19
2.3. EXPERIMENT 2 23
2.4. EXPERIMENT 3 26
2.5 GENERAL DISCUSSION 29
3. FACE IMAGERY IS BASED ON FEATURAL REPRESENTATIONS 32
3.1 INTRODUCTION 32
3.2 EXPERIMENT 4 34
3.3 DISCUSSION 38
4. THE THATCHER ILLUSION: ROTATING THE VIEWER INSTEAD OF THE PICTURE 41
4.1 INTRODUCTION 41
4.2 EXPERIMENT 5 44
4.3 DISCUSSION 48
5. IS THE THATCHER ILLUSION RESTRICTED TO FACES? 51
5.1 INTRODUCTION 51
5.2 EXPERIMENT 6 52
5.3 DISCUSSION 54
6. ARE FEATURAL AND CONFIGURAL FACE PROCESSING STRATEGIES DISSOCIABLE? EVIDENCE FROM AN FMRI STUDY 56
6.1 INTRODUCTION 56
6.2 EXPERIMENT 7 58
6.3 DISCUSSION 60
7. SYNOPSIS: WHERE DOES ALL THIS LEAVE US? 64
7.1 INTRODUCTION 64
7.2 THE DEFEAT OF THE HOLISTIC HYPOTHESIS 64
7.3 TOWARDS A NEW MODEL OF FACE PERCEPTION 65
7.4 LIMITATIONS 67
7.5 CONCLUSION 68
8. GLOSSARY 69
9. REFERENCES 71
ACKNOWLEDGEMENTS 81
CURRICULUM VITAE 82

1. Face processing: A topic of wide interest

Abstract

The following chapter reviews the literature in the field of face processing. First, different hypotheses of face processing are discussed: the holistic hypothesis, schema hypothesis and configural-featural hypothesis. In a further paragraph literature on the face inversion effect is reviewed. Then, higher cognitive functions of face processing are addressed, focussing specifically on mental imagery and mental rotation of faces. Finally, evidence from brain-imaging studies on face perception is discussed.

1.1. Introduction

Face perception is probably the most highly developed visual skill in human beings. From a very early age infants prefer to look at faces than to other objects. Already 30 minutes after birth infants track a moving face farther than other moving patterns of comparable complexity (Johnson, Dziurawiec, Ellis, & Morton, 1991). Small children will soon learn to discriminate between the familiar faces of their parents and faces of strangers. While the parents' faces are associated with safety and cosiness, strangers may evoke a fearful reaction. The fact that face perception is so highly developed is not surprising, as faces are a biologically relevant stimulus class. Processing information contained in faces is a pivotal social skill. Every time we encounter a face we essentially process information that tells us whether the face is familiar, what emotion the person conveys and even where his or her attention is focussed on. And these processes can be initiated for an almost infinite number of faces. Indeed, it is surprising how well human beings can recognize and discriminate innumerable different faces. According to a study by Bahrick, Bahrick, and Wittlinger (1975) not only face discrimination, but also the long-term memory of faces is very accurate; they found that faces can be recognized with a 90% precision after more than 50 years. The biological relevance of faces for human beings makes faces a very important object category and it can be said, at least for adults, that they are real experts. This expertise helps us to overcome a further characteristic of face perception: Faces are typically recognized at the exemplar-specific level. In contrast to objects, which we often recognize at the basic level (e.g. "chair" or "bottle"), faces are generally recognized at the most extreme subordinate level (e.g. "John" or "Paul"). Humans are highly expert at discriminating between individual faces.

Apart from their biological relevance, empirical data have suggested that face recognition is somehow "special". It has been claimed that faces are processed differently than other object classes. But what exactly is so "special" about face perception? Some researchers suggested that faces are processed holistically. They understand faces as a class of stimuli that are encoded and recognized as whole templates, without representing parts explicitly (Farah, Wilson, Drain, & Tanaka, 1998; Murray, 2004; Palermo & Rhodes, 2002;

Robbins & McKone, 2003; Tanaka & Farah, 1993). In contrast, other objects categories such as furniture, houses or tools have been thought of as being processed part-based, each part of an object being stored separately (e.g., Biederman, 1987; Marr, 1982). The holistic hypothesis is discussed in section 1.2. Another aspect that seems unique for faces is the great expertise human beings have with faces. A number of authors claimed that expertise with another object class (e.g. dog experts) show similar processing characteristics for objects from their field of expertise as for faces (e.g., Diamond & Carey, 1986). This assumption is reviewed in section 1.3. Finally, some authors adopt a dual-mode view, claiming that configural and featural information is processed following two distinct pathways (e.g., Bartlett, Searcy, & Abdi, 2003; Cabeza & Kato, 2000; Schwaninger, Lobmaier, & Collishaw, 2002). The featural-configural hypothesis is discussed in section 1.4 (see also chapters 2, 3, and 5).

Another peculiarity is that faces are a lot more sensitive to inversion than items of other object classes. Recognition of upside-down faces drops significantly, and also two simultaneously presented inverted faces can only be matched with great difficulty. This face inversion effect (FIE) is a robust phenomenon, it has been found when participants name, classify, or match photographs or drawings of faces (for an overview see Valentine, 1988). The face inversion effect is addressed in section 1.5 (see also Chapter 4 and 5). Closely related to the FIE and a possible explanation for the difficulty to recognize upside-down faces is the presumption, that recognition of disoriented visual stimuli requires a mental rotation of the stimulus. For disoriented faces, the spatial transformation of all features and configurations overtaxes the capacity of the underlying mechanism (c.f. Rock, 1973). Therefore it is difficult to mentally visualize what an inverted face would look like right side up. In section 1.6 another cognitive skill is discussed, namely mental visual imagery of faces. There is evidence that perception and imagery at least partly share same mechanisms. Studies showing differences and similarities between face perception and face imagery are reviewed (see also chapter 3).

In section 1.7 some characteristics of face processing are explained and illustrated with evidence from neuroscience. Studies with neurological patients have shown that lesions in a certain area of the temporal lobe lead to selectively impaired face recognition, while the ability to recognise objects of similar visual complexity was unaffected. This was interpreted as implication that faces are processed in a different brain area than objects. Indeed, neuro-imaging studies have confirmed that faces evoke selectively more activation in an area in the fusiform gyrus consequently termed the fusiform face area (FFA, e.g., Kanwisher, McDermott, & Chun, 1997; Kanwisher, Stanley, & Harris, 1999). Neurophysiological findings further substantiated this proposition. Single unit recordings in the macaque brain identified neurons in the superior temporal sulcus of the macaque brain that responded selectively to faces (e.g., Desimone, 1991; Perrett, Hietanen, Oram, & Benson, 1992). Perrett and colleagues (1992) found cells that selectively respond to different aspects of faces, such as different views of the face and head, or cells that discriminate familiar from unfamiliar faces. However, given the amount of different object classes it seems unlikely that there is a dedicated brain region for every object class. This observation can be replied to with the assumption of distributed neural networks as proposed for example by Ishai, Ungerleider, Martin, Schouten, & Haxby (1999). In section 1.8 conclusions are drawn from the previous sections and merged together.

1.2. Holistic hypothesis of face perception

The holistic hypothesis of face perception claims that faces are represented and recognized as undifferentiated wholes (cf. Tanaka & Farah, 2003). In the early 20th century this type of visual apprehension was characterized as “gestalt”, where the perception of the whole takes precedence over the sum of its constituent parts. Various interpretations of holistic face processing have been suggested. The pure holistic view claims that faces are represented as whole templates and that facial parts (or features, components) are not explicitly represented (Tanaka & Farah, 1993; see also Farah, Tanaka, & Drain, 1995). According to such a view, faces are stored as unparsed perceptual wholes without explicitly storing information of the individual parts. A simple analogy of such a holistic face representation would be a bitmap, in which only colour values of points are specified without providing information about which point belongs to the eyes or nose. The bitmap indeed contains eyes and nose, but it does not represent them explicitly. A more moderate view of the term “holistic” is held by Maurer and colleagues (Maurer, Le Grand, & Mondloch, 2002), who claim that holistic processing is one type of configural processing in which the features are glued together into a gestalt. According to Maurer and colleagues, configural processing refers “to any phenomenon that involves perceiving relations among the features of a stimulus such as a face” (Maurer et al., 2002, p. 255). This definition is very similar to what has been defined as configural information (see below).

A line of studies have been interpreted in favour of the holistic hypothesis. For example, Tanaka and Farah (1993) trained participants in recognizing upright faces. In the experimental phase, two faces which differed either in the shape of the eyes, nose or mouth were simultaneously presented. In a second experimental condition the eyes, nose or mouth were presented in isolation, that is without the facial context. Participants had to judge which of these faces appeared in the training phase. The authors found that it was more difficult to recognize a part of a previously learnt face when it was presented in isolation than when it was embedded in the facial context. This difficulty to recognize isolated parts was interpreted in favour of a holistic view of face processing. Parts, the authors concluded, are therefore not explicitly represented. Similarly, Farah et al. (1998) found that faces are represented more holistically in immediate perceptual memory and during perception as compared to inverted faces, houses and words. Yet they argue that faces are not “special” because holistic representation is confined to faces, but because face recognition is the extreme end of a continuum of part-based to more holistic representation. Tanaka and Sengco (1997) hold a slightly different view of holistic face processing. Similar to Maurer et al. (2002) they claim that featural information and configural information are combined into holistic face representations. Whereas Tanaka and Farah (1993) and Farah et al. (1995) claimed that faces are represented as unparsed wholes without any representations of parts, Tanaka and Sengco (1997) imply that featural and configural information are first represented separately before they are integrated into a holistic representation (see also Rhodes, Brake, & Atkinson, 1993).

1.3. Schema hypothesis of face processing

Another hypothesis has been suggested by Goldstein and Chance (1980). A schema is understood as an organizing mechanism for both information input and output. According to the schema theory, this mechanism develops and improves according to experiences gained through interaction with incoming stimuli. Such an improvement is achieved at the expense of flexibility; with growing number of processed “normal” faces the ability to handle variational faces declines. With this theory Goldstein and Chance (1980) explain the observation, that faces from a foreign racial group seem to resemble each other much more than faces of the own race. This other race effect emerges through constant exposure to faces of our own racial group: By means of interaction between incoming stimuli the face schema is modified and is able to efficiently process a wider range of stimuli. But this enhancement goes with reduced flexibility towards deviant faces, resulting in hampered ability to process faces of other racial groups. The term schema rigidity has been suggested by Goldstein (1975) to describe the reduction of flexibility resulting from overlearning face stimuli.

The schema hypothesis may explain the observation that faces seem to be processed holistically (see Diamond & Carey, 1986; Rhodes, Tan, Brake, & Taylor, 1989; Tanaka, 2001; see also Gauthier & Tarr, 1997). Diamond and Carey (1986) provided direct evidence that expertise might be responsible for “face specific” processing. They tested dog experts on upright and upside-down dog pictures and found a similar inversion effect as for pictures of human faces, but only for dogs of the breed for which the participants were experts. Insofar face processing differs from other object processing in the expertise we have with them. As a further consequence of expertise a qualitative shift in processing can be observed. For instance in a study by Tanaka and Taylor (1991), bird experts were as fast to recognize objects at the subordinate level (“robin”) as they were at the basic level (“bird”). Non-experts were consistently faster on basic-level discrimination as compared to subordinate-level discriminations. Similarly, because humans are face experts, judgments of face identity (subordinate level) are as fast as judgments that are more categorical, for instance gender (Tanaka & Taylor, 1991; see also Gauthier & Tarr, 1997).

Mondloch, Le Grand, and Maurer (2002) showed that expertise for face recognition takes many years to develop. Especially configural processing (see section 1.4) seems to be fully developed only after the age of 10 years. In a same-different task Mondloch, et al. (2002) presented adults and children of 6, 8, and 10 years of age pairs of upright and inverted faces. Either the faces were the same, or they differed either in the shape of eyes and mouth (featural set), in the distance between the eyes and the mouth (spacing set), or in the shape of the external contours (contour set). Children generally made more mistakes in the spacing set, giving evidence that their configural processing system has not yet fully developed.

1.4. Featural-configural hypothesis of face perception

Recent studies have shown that a pure holistic view of face perception does not hold. Representations of featural and configural information are two important sources of information in face recognition (Bartlett et al., 2003; Cabeza & Kato, 2000; Schwaninger et al., 2002). Featural information is referred to as the information contained in the components (or parts) of a face (i.e., eyes, nose, mouth etc); configural information is understood as the spatial relationship between the features (e.g., Bruce, 1988). The spatial interrelationship of facial features was further differentiated by Diamond and Carey (1986) who distinguished first-order and second-order relational information. First-order relational information refers to the basic arrangement of the parts (e.g., the nose lies between the eyes), whereas second-order relational information means the exact metric distances between the features.

Already Sergent (1984) suggested that faces have featural and configural properties and proposed that these two properties are processed following two distinct processing strategies. It is suggested that these processing strategies are not mutually exclusive and can unfold simultaneously. Sergent (1984) used a two-choice “same-different” judgement task and a dissimilarity judgement task on the same set of stimuli to examine the nature of configural and featural processing. In each task a face pair was presented in which the two faces were either the same, or differed in 1 to 3 different features (eye region, internal spacing, or inferior contour. Sergent found that the reaction times decreased linearly as the number of differences between the faces increased. This result suggests that the facial features combine additively and are processed independently of one another. However, not all her feature manipulations were equally salient: faces differing in contour were compared faster than faces differing in eyes or internal spacing. Similarly, faces differing in contour were judged as less similar than faces that differed in other dimensions. It has to be noted here that the feature manipulations used by Sergent (1984) do not involve featural information as defined above. Her manipulations included clear configural information, such as for example internal spacing. Insofar Sergent’s paradigm did not enable separate investigation of featural and configural information.

Most studies on configural and featural face perception suggest that configural processing plays a dominant role over featural processing (e.g., Bartlett et al., 2003; Cabeza & Kato, 2000; Diamond & Carey, 1986; Haig, 1984; Leder & Bruce, 2000; Schwaninger et al., 2002). Yet, this sensitivity to configural information is not face-specific. Gauthier and Tarr (1997) found evidence that this sensitivity to configural changes might be explained by a more general recognition mechanism fine-tuned by experience with homogeneous stimuli. They used “greebles” as stimuli in a series of experiments. Greebles are a computer-generated novel class of objects which are comparable to faces, as all exemplars share the same number of parts in the same configuration. Gauthier and Tarr (1997) extensively trained participants on a set of greebles and found a shift from featural to configural processing. Furthermore, Gauthier, Tarr, Anderson, Skudlarski, and Gore (1999) found that, with growing expertise with greebles, fMRI activation in the right fusiform face area (FFA) resembles the pattern activity of upright faces (see section 1.7 below). From this standpoint predominance of configural processing is not per se face specific, but is modulated by expertise. Similar effects have been found for bird and car experts (Gauthier, Skudlarski, Gore, & Anderson, 2000), and dog experts (Diamond & Carey, 1986).

Studies investigating configural and featural face information predominantly used stimuli where the configuration was changed (e.g. Haig, 1984; Macho & Leder, 1998; Searcy & Bartlett, 1996; Sergent, 1984; Tanaka & Sengco, 1997) or where the features were changed (e.g. Farah et al., 1998; Searcy & Bartlett, 1996; Sergent, 1984; Tanaka & Farah, 1993). Such manipulations are problematic, as configural changes may also involve featural changes and vice versa (Rakover, 2002; Sergent, 1984). For example stretching the inter-eye distance may result in the bridge of the nose appearing wider. Schwaninger, Lobmaier, & Collishaw (2002) circumvented this problem by using scrambled and blurred faces. Scrambled faces contain only information of the features, whereas, if sufficiently blurred, all featural information is lost in blurred faces and isolated configural information remains. To test whether all configural information is destroyed in the scrambled faces and vice versa, the faces can be both blurred and scrambled at the same time. If the faces are sufficiently blurred, such faces will no longer be recognized (Schwaninger et al., 2002).

A configural representation is qualitatively different than a featural representation, as it contains metric distances and spatial relations. Diamond and Carey (1986) suggested that with growing expertise items are represented more configurally. As human beings we are experts in face recognition, therefore configural information plays a larger role than featural information. Additionally, familiarity of a face plays an important role. Buttle and Raymond (2003) found that familiar faces are processed more efficiently than recently learned faces. They claimed that this is due to a more configural mode of analysis in familiar faces (Buttle & Raymond, 2003).

1.5. The face inversion effect

A characteristic which seems to be distinctive for faces is what has been described as the face inversion effect (FIE), first reported by Yin (1969). He found a much poorer performance for recognizing inverted faces as compared to other inverted objects, such as airplanes, houses, or stick figures of men in motion (Yin, 1969). The FIE is a robust phenomenon, it has been found when participants name, classify, or match photographs or drawings of faces (for an overview see Valentine, 1988). An impressive demonstration of how inverted faces are difficult to process was provided by Thompson (1980). He took a photograph of the former British prime minister Margaret Thatcher and inverted eyes and mouth with respect to the whole face. Such a face looks extremely grotesque when viewed right-side up, but loses this grotesqueness when the face is inverted (Thompson, 1980). This effect is now commonly referred to as Thatcher illusion. Numerous studies have been concerned with this phenomenon since then (e.g., Lewis, 2001; Rakover, 1999; Sturzel & Spillmann, 2000; Valentine & Bruce, 1985). Sturzel and Spillmann (2000) gradually turned different thatcherized faces from 0° through 180° and asked participants to report when the face switched from pleasant to grotesque, or vice versa. They found a relatively narrow changeover-zone between 97.2° and 118.3° where the change of expression occurred. Sturzel and Spillmann suggested that the striking change may be based on the step-tuning properties of hypothetical face neurons, rather than a gradual tuning curve. According to Sturzel and Spillmann face neurons respond best to faces in a tuning width of ±100° relative to the vertical. They claim that these face neurons may also respond to inverted faces, but inappropriately. In contrast, Lewis (2001) reported a gradual loss of configural information the further a face is turned away from upright. He recorded the reaction times of 40 participants while they discriminated thatcherized from normal faces which were presented in 10 different orientations. Based on these two studies, however, the nature of the dependence on the rotation-angle still remains equivocal.

Why is it so difficult to process an inverted face? An explanation for the face inversion effect on the basis of the holistic and schema hypotheses is that faces are mono-oriented and clearly have an upright orientation. We rarely see faces upside-down and we may therefore only have an upright representation of faces. Furthermore, Farah et al. (1995) claim that perception of holistically represented complex patterns is orientation sensitive, and thus inversion makes encoding conditions difficult. Further evidence that holistic processing is hampered in inverted faces has been found when the top half of a face was aligned with the bottom half of another face. When presented upright, adults found it very difficult to identify the top half of such a composite face, as they seemed to perceive the face as a whole gestalt. When the face was turned upside-down, or when the two halves were misaligned, performance improved, because the face did no longer appear as a whole (Hole, 1994; Stevenage, 1995; Young, Hellawell, & Hay, 1987).

An alternative answer is the mental rotation hypothesis suggested by Rock (1973), based on the featural-configural hypothesis: When we encounter inverted faces, we mentally rotate the input picture. In inverted faces "there is a whole set of component figures and figural relationships have to be corrected, and it is not possible to succeed in visualizing simultaneously how each of these would look were it to be egocentrically upright" (Rock, 1973, p. 60f). With the mental rotation hypothesis, Irving Rock provided a possible explanation for the Thatcher illusion seven years before Thompson first reported this effect. A slightly different explanation for the Thatcher illusion is provided by Rakover (1999). He interprets the strangeness of upright thatcherized faces as a result of our inability to grasp the eyes as locally inverted in the pattern of the upright whole face. When inverted, the thatcherized face is perceived as an inverted face and the eyes are perceived in this frame as they are: upright eyes. But in this inverted case, the whole face is perceptually not as dominant as the individual features (because inversion disrupts perception of the whole face).

Meanwhile, a widely accepted explanation for the FIE is that inversion substantially impairs processing of spatial-relational information. According to Leder and Bruce (1998) it is the disruption of configural information instead of holistic information in inverted faces which accounts for the face inversion effect (see also Boutsen & Humphreys, 2003; Leder, Candrian, Huber, & Bruce, 2001; Maurer et al., 2002; Rock, 1973; Searcy & Bartlett, 1996; Sergent, 1984; Yin, 1969; for a review see Schwaninger, Carbon, & Leder, 2003). In their recent review, Rossion and Gauthier (2002) suggested that "most if not all of the decrease in face discrimination performance caused by inversion can be accounted for by the disruption in inverted faces of the processing of the local spatial relationships between features" (Rossion & Gauthier, 2002, p. 55; see also Leder & Bruce, 2000). As reported above, it is claimed that perception of faces relies predominantly on configural information; therefore the inversion effect is so pronounced with faces. This leads to a slightly revised formulation of Rakover's (1999) explanation of the Thatcher illusion. In my opinion, the inverted thatcherized face does not seem as strange, because such a face can no longer be processed configurally. Instead, we have to rely on featural information, which is not as sensitive to inversion. We clearly perceive such a face as inverted, but miss that the eyes and mouth are in fact upright. As a consequence we fail to notice the grotesqueness of such a face. In the upright thatcherized face however, featural and configural processes act jointly and the grotesqueness of the face is apparent.

Studies aiming to investigate configural and featural processes have often made use of the disproportionate sensibility of featural and configural information to inversion. Whenever an inversion effect was found authors concluded that configural processing was involved. For example, Rhodes and colleagues (Rhodes et al., 1989) showed that “own race” faces exhibit a larger FIE than faces of the “other race”. They assumed that “own race” faces constitute high expertise, whereas “other race” faces signify low expertise. From their findings they conclude that expertise is associated with greater use of configural information in faces. Diamond and Carey (1986) found that dog-show judges and dog breeders showed a comparable inversion effect for dogs as for faces, but only for dogs of the breed of which they were experts. This finding led the authors to the conclusion that with growing expertise a more configural processing mode is adopted. Some findings that familiar faces are processed more configurally than unfamiliar faces are also based on the FIE. For example, Buttle and Raymond (2003) successively showed their participants two pairs of faces. In the second pair one of the faces was changed. The task was to detect the changed face; no explicit recognition or naming was involved. Performance was significantly better if the changed face was famous. However, this advantage for familiar faces was abolished, when the faces were inverted. The authors concluded from this result that familiar face processing predominantly activates a configural mode of analysis. The conclusions of Mondloch et al. (2002) that configural processing takes longer to develop was partly drawn from results involving the inversion effect. For adults and 10-year-olds the FIE was larger in pairs differing in the distances between the features than in pairs differing in the shape of the features or outer contours. However, the size of the inversion effect did not differ between the conditions in the younger children. The authors presume that the six and eight-year-olds were less able to take advantage of configural information in upright faces (Mondloch, Le Grand, & Maurer, 2002). Similar results have been found by Freire and Lee (2001).

It is important to note, however, that although configural information seems highly sensitive to inversion, an inversion effect does not necessarily mean configural processing. For example, Schwaninger, Lobmaier, and Fischer (2005) found an inversion effect for gaze perception. But this effect was found for whole faces and also when only the eyes were presented in isolation. This finding led the authors to suggest that the gaze inversion effect is probably the result of perceptual learning (Schwaninger, Lobmaier, & Fischer, 2005).

1.6. Mental imagery of faces

Visual images are vision related experiences that are triggered when we recall information from long term memory. Visual imagery is referred to the processes involved in generating and examining such images. Visual imagery is a basic form of cognition and plays a central role in many human activities and according to Kosslyn (1994) imagery is likely to be one of the first higher cognitive functions firmly rooted in the human brain. Even in pre-scientific ages imagery has been the concern of many philosophers, back as far as Aristotle.

Almost everybody knows what the Eiffel Tower looks like and to some degree will report some kind of visual experience when thinking of it. This affinity of visual perception and visual imagery has lead to a great dispute of whether and to which extent imagery and perception are alike (for an overview, see Kosslyn, 1994). In his model, Kosslyn (1994) suggests three basic components, (1) the visual buffer, (2) long-term stored representations and (3) image-processing operations. The visual buffer is a spatially organized array within the visual system and constitutes the medium for visual images. Visual images are generated by retrieving information from long-term memory and can be inspected and transformed within the visual buffer. Visual imagery is initiated by active processes which generate images in the visual buffer from the stored descriptions in long-term memory. These processing operations are also responsible for inspection and transformation of visual images. According to Kosslyn (1994) the difference between perception and imagery lies in the generating source of visual images in the visual buffer. While in perception the visual buffer is activated "bottom-up" by visual input, it is triggered "top-down" in imagery.

Visual mental imagery and visual perception have been reported to have common properties and even share neural substrates (e.g. Ishai, Ungerleider, & Haxby, 2000; Ishai, Haxby, & Ungerleider, 2002; Kosslyn, 1994; O'Craven & Kanwisher, 2000). Both can be used for priming, albeit priming from visual imagery is weaker (Michelon & Zacks, 2003). Parallel to the "what" and "where" system in perception (c.f. Ungerleider & Mishkin, 1982), a dissociation of visual-object and visual-spatial imagery could be demonstrated in studies of patients with brain damage (Levine, Warach, & Farah, 1985).

Because face perception seems to be a special form of visual perception, face imagery suggests itself to be a somewhat special kind of visual imagery. However, face imagery can hardly be separated from person-specific information that is not purely visual (character, sympathy, voice etc.). Here, only visual aspects of face imagery are addressed.

1.7. Evidence from neuro-imaging studies and neurophysiology

Numerous brain imaging studies on face perception have already been conducted. Many authors were in search of a region that is specialized on face processing. Indeed, a large majority of them found activation in the right fusiform gyrus while processing faces compared to objects (e.g., Ishai et al., 1999; Kanwisher et al., 1997; Kanwisher et al., 1999; Rhodes, Byatt, Michie, & Puce, 2004). Such findings suggest that the right fusiform gyrus is specialized at least for some aspects of face perception and this region of the brain has therefore been termed fusiform face area (FFA). In support for such a specialized system for face processing, Perrett and colleagues (Perrett et al., 1991) found cells in the STS selectively responsive to faces. More so, while some cell populations responded maximally to upright front-view faces, other cell populations responded more to profile view faces.

Other authors suggested that the FFA is not a region specialized for faces, but a region that reflects expertise with a certain object class. Evidence for this idea was found in studies with greebles, where with growing expertise with greebles fMRI activation in the right fusiform face area (FFA) resembled the pattern activity of upright faces (Gauthier et al., 1999). Similar effects have been found for bird and car experts (Gauthier et al., 2000). However, Rhodes et al. (2004) used faces, Lepidoptera, and common objects and tested novice participants in a free viewing and an individuation task. In a second experiment they tested Lepidoptera experts. In both groups the FFA responded grater to faces than to Lepidoptera or common objects, supporting the face-specifity hypothesis of the FFA.

Finally, a number of authors postulate a distributed neural system for face processing (Haxby et al., 2001; Haxby, Hoffman, & Gobbini, 2000; Hoffman & Haxby, 2000; Ishai, Ungerleider, Martin, & Haxby, 2000; Ishai et al., 1999). These authors argue that although it may be possible that there are dedicated neural modules for certain biologically relevant objects such as faces, which have emerged through evolution, it seems very unlikely that there are modules for all object categories. Instead they suggest that the representations of objects in the ventral temporal cortex are more widely distributed. Indeed, all objects activate a broad expanse of ventral temporal cortex, albeit to varying degrees, suggesting that the representation of objects in this cortex may be feature rather than object based. Ishai et al. (1999) found three adjacent regions that responded differentially to houses, faces and chairs. Indeed, these findings could be interpreted as evidence for separate modules. However, all three regions also responded significantly to other object classes. The authors conclude that the representations of objects are distributed across ventral temporal cortex and are not restricted to category specific anatomically segregated modules.

Can behavioural findings of two separate routes to face processing be replicated and substantiated using neuro-imaging? Only a few studies investigating the issue of two separate mechanisms for featural and configural processing have been accomplished so far. For example, Rossion et al. (2000) found more activation in the left FFA if subjects focused their attention on particular features of faces. If, on the other hand, participants relied more on configural mechanisms to process faces the responses were larger in the right FFA. Part-based processing therefore reduced face-specific activity in the right fusiform gyrus. Haxby and colleagues (Haxby et al., 2000) suggested that a region of the inferior occipital gyrus may be involved in the perception of facial parts. Perrett, Rolls, and Caan (1982) recorded single cells in the STS and found neurons that were selectively responsive to parts (i.e. eyes, mouth or hair). Different cells responded to different features, or subsets of features. The superior temporal lobe further seems to be involved in processing changeable aspects of faces (Haxby, et al. 2000) and spatial relations (Leube et al., 2003). Leube et al. (2003) found more activation in the right superior temporal sulcus and right insula for upright faces than for inverted faces and concluded that these regions process configural properties. The authors argued that in inverted faces configural processing is hampered, hence the reduced activation. Sagiv and Bentin (2001) compared the N170 event-related potential (ERP) component triggered by faces of different authenticity (photographs, painted portraits, sketches of faces, and schematic faces) and found that N170 did not distinguish between different face types when presented upright. However, the N170 was enhanced for inverted natural faces and reduced for inverted schematic faces. The authors concluded from this finding that early face processing is subserved by a multiple-component neural system processing both configural and featural information. They claim that the relative involvement of the two processes is determined by whether the presented face activates a holistic perception process or an analysis of the features (Sagiv & Bentin, 2001).

As reported above, Rossion et al. (2000) suggest a left-lateralization for part-processing. However, this evidence comes from studies where participants had to focus their attention on certain features of the face. This does not necessarily mean that representations of features can be located in the left hemisphere. Also larger responses in the right fusiform face area when participants rely more on configural information does not necessarily mean that configural representations can be located in the right fusiform gyrus. A lateralization of brain activation may simply reflect different strategies in the same face recognition process. Indeed, no dissociated systems are expected on the basis of a distributed neural network (Haxby et al., 2001; Haxby et al., 2000; Ishai, Ungerleider, Martin et al., 2000; Ishai et al., 1999). Instead, the difference between configural and featural processing is expected to be the result of quantitative activation differences within the network.

There is a scarce number of brain imaging studies investigating whether face perception shares similar underlying neural networks as face imagery. These studies suggest that visual imagery evokes – at least partly - similar activation as when the faces are in fact perceived (Farah, Peronnet, Gonon, & Giard, 1988; Ishai et al., 2002; Ishai, Ungerleider, & Haxby, 2000). In an fMRI study, Ishai and colleagues (Ishai, Ungerleider, & Haxby, 2000) found content-related activation in extra-striate cortex and ventral temporal cortex when the participants visually imagined faces, houses, and chairs. Further evidence that imagery and perception of faces underlie similar neural mechanisms comes from case studies with prosopagnosic patients, which revealed that an impairment of face recognition is often accompanied with the disability to mentally visualize faces (Charcot & Bernard, 1883; Young, Humphreys, Riddoch, Hellawell, & de Haan, 1994; Young & Van De Wal, 1996). Some reports, however, have described prosopagnosic patients with intact face imagery (e.g., Bodamer, 1947; Pallis, 1955).

Some studies even found a dissociation between imagery of configural and featural information. For example, Ishai et al. (2002) found more activation in the right intraparietal sulcus (IPS) and the inferior frontal gyrus (IFG) when participants attended to features of the imagined faces as compared to a more holistic image of the faces. They interpret this increase of activation as a result of attention, not as activated featural representations.

1.8. Conclusion

It is evident from the large number of studies that face processing is a topic of wide interest. A question conveying most research on face processing is whether faces are "special". Indeed, some findings suggest that faces are processed differently than other object classes. In the 1990's scientists were debating whether faces were processed holistically. Behavioural findings were interpreted such that faces are processed as perceptual wholes without representing the individual parts. What is special about faces according to these authors is that whereas object recognition depends on the decomposition of the object into its constituent parts, faces are represented and recognized as undifferentiated wholes (e.g. Tanaka & Farah, 2003). Alternatively, some authors suggested that every visual object is processed both configurally and part-based, the speciality of face processing being that in faces configural information plays a relatively more important role than featural information.

Still other authors alleged that the speciality of faces merely lies in the experience we have with recognizing and processing faces. Indeed, experts in the field of a certain object class showed similar recognition characteristics with objects from their field of expertise as with faces (e.g. Diamond & Carey, 1986; Gauthier & Tarr, 1997).

All these hypotheses can be accounted for face processing, and indeed, the different hypotheses partly overlap. But of all, the featural-configural hypothesis can explain face processing most aptly. Findings seemingly suggesting that faces are processed holistically can just as well (or even better) be explained by properties of configural processing. For example, Tanaka and Farah (1993) found that parts of faces were recognized better when presented in the context of whole face can be explained better by the featural-configural hypothesis. Given that faces are predominantly processed on the basis of configural information, it is not surprising that the individual parts are recognized more accurately when they are presented in the configural context. Allowedly, various interpretations of the term holistic processing have been suggested and holistic and configural processing has often been used interchangeably. But to conclude from such results that parts are not explicitly processed and represented seems a bit overhasty. Many arguments that were interpreted in favour of the holistic hypothesis seem to be the result of misapprehension of the definitions of holistic, featural and configural processing. For instance, Tanaka and Farah (2003) misleadingly interpreted the findings of Sergent (1984) as evidence against the "featural position". Sergent (1984) found that RTs to differences of chin contour was faster than differences in internal spacing. Moreover, differences in chin contour and internal spacing produced even shorter RTs than differences in chin contour alone. According to Tanaka and Farah (2003) the "featural position" demands that the time to make different responses when faces vary by two features should never be faster than the time to make a different response when faces differed by the most salient feature. This objection does not hold, because differences in internal spacing are by definition a property of configural information. The fact that Sergent did not find these effects in inverted faces further substantiates this misunderstanding.

Further evidence in favour of the featural-configural hypothesis comes from studies of the face inversion effect (FIE). The FIE seems to be a consequence not of orientation sensitivity of holistic representations, but of the disproportionate orientation sensitivity of featural and configural information (e.g., Leder & Bruce, 1998; 2000; Searcy & Bartlett, 1996). Furthermore, studies seemingly supporting a holistic view of face perception can just as well be interpreted in favour of the featural configural hypothesis. For example Farah et al. (1995) argued that faces are orientation sensitive because face perception is holistic and the perception of holistically represented complex patterns is orientation sensitive. This finding however can be explained by differential sensitivity to inversion of features and configural information. Because, as has been shown in most studies on face perception, faces seem to be processed mainly on the basis of configural information, inverted faces are disproportionately more difficult to recognize than inverted exemplars of other object classes. Finally, the expertise effect can be explained by the relative more important role of configural information in exemplars in the field of expertise.

The featural-configural hypothesis implies that featural and configural information can be processed independently. Indeed, there is empirical evidence to assume two separate routes to face perception. In various behavioural studies faces could be recognized both by the features and by the configuration (e.g., Collishaw & Hole, 2002; Schwaninger et al., 2002). Further, once again findings from inversion studies support the assumption of two separate processing mechanisms. Thus, inversion hampers configural processing whereas features can still be processed without mentionable constraints. Additional evidence for two separate processing mechanisms comes from developmental studies. Mondloch, Le Grand & Maurer (2002) found that configural face processing develops more slowly than featural face processing. In a same-different task they presented adults and children of 6, 8 and 10 year of age with upright and inverted face pairs. Either the pairs were the same, or they differed in their features or configuration. Children generally made more mistakes in configural pairs compared to adults, giving evidence that their configural processing system has not yet fully developed. Furthermore, 6- and 8-year-old children showed a comparable inversion effect for featural and configural pairs whereas 10-year-olds showed a larger inversion effect for configural pairs (see also Freire & Lee, 2001). Finally, supporting the idea of two separate pathways, configural and featural processing could be dissociated anatomically (e.g. Rossion et al., 2000; Bartlett et al., 2003).

The emergence of brain-imaging techniques such as fMRI led the dispute to the question of how faces are anatomically represented in the brain. Indeed, brain-imaging studies revealed an area in the fusiform gyrus, consequently called the fusiform face area (FFA), which is supposedly specialized on faces. Such a specialized face area finds support in studies with patients suffering from prosopagnosia. These people have severe difficulties in recognizing familiar faces, whereas their performance in recognizing items of other object classes remains unimpaired. This deficiency to recognize faces can be ascribed to lesions fusiform gyrus, namely in the FFA. Further support for a specialized system is provided by studies of single cell recordings in non-human primates. For example, Perrett and colleagues found cells in the macaque STS that were selective for faces. Whereas such findings can be interpreted in favour of a specialized neural system for face perception, there is evidence that questions a special module for face processing. For instance, if we assume a face module a coherent implication would presume a specialized system for living objects such as animals or trees, or inanimate objects such as chairs, bottles, or different tools. Such a proposition would overcharge the capacity of our brains. Further, not only the FFA, but also other brain regions are activated by faces. For example, Kanwisher et al. (1997) also found a greater activation for faces than objects in the middle temporal gyrus and in the STS. Further still, expertise with a certain object class evokes activation in the FFA, suggesting that the FFA is not selective for faces (e.g. Gauthier, etc.). In addition, Joseph and Gathers (2002) found that the FFA was also activated by natural and manufactured objects. Such objections can be replied to by postulating distributed neural systems for object recognition (e.g., Haxby et al., 2001; Haxby et al., 2000; Ishai, Ungerleider, Martin et al., 2000; Ishai et al., 1999).

In the following chapters the assumption is made that features and configuration are two important modes of information in face perception. Employing a dual-code view of face perception, I assume that configural and featural representations are formed when we encounter a face and are activated when we have to recognize a face. It is presumed that these representations can be activated bottom-up and top-down, in perception and in imagery. In the following chapters the nature of these representations and their interrelationship are studied. I use scrambled and blurred faces to separately investigate the role of featural and configural face information. In the scrambled faces eyes, nose and mouth were cut out and were arranged so that no feature came to lie in its correct categorical relation to its neighbouring part. Thus, the detail information of each facial part was maintained while disrupting the spatial relation to the other parts. For configural faces, the stimuli were blurred using a low-pass Gaussian filter. By removing higher spatial frequencies from the spectrum present in a picture of a face, the overall configuration of the face remains intact, while the details of the features are unspecified. Chapter 2 ascertains following questions: When are featural and configural representations formed? Are they formed as soon as we see a face? Or do we have to see a face longer and repeatedly before we form featural and configural representations? Is there an advantage for one kind of representation (e.g., do we memorize faces better when storing individual features or when storing the configuration of the features)? The study reported in chapter 3 investigates whether configural and featural information can be differentiated in mental imagery. Specifically I ask whether mental imagery primes featural or configural information, or both. In chapter 4 a study is reported where the influence of the gravitational reference frame on the Thatcher illusion investigated. Chapter 5 addresses the question whether the Thatcher illusion is restricted to faces. And finally, in chapter 6 a neuro-imaging study is reported scrutinizing whether featural and configural processes can be dissociated using fMRI. The concluding chapter 7 discusses the results of chapters 2 – 6 and reconciles them with previous findings reported in chapter 1.

2. Perception of Novel Faces: The Parts Have It!

Abstract

It has been suggested that, as a result of expertise, configural information plays a predominant role in face processing. This idea was investigated using novel and learned faces. In Experiment 1 sixteen participants matched two subsequently presented blurred or scrambled faces which could be either upright or inverted in a sequential same-different matching task. By means of blurring, featural information is hampered, scrambling disrupts configural information. Each face was unfamiliar to the participant and was presented for 1000 ms. An ANOVA on the *d'* values revealed a significant advantage for scrambled faces. In Experiment 2 fourteen participants were tested using the same design, except that the second face was always intact. Again, the ANOVA on the *d'* values revealed a significant advantage for scrambled faces. In Experiment 3 half of the faces were extensively learned in a familiarization block. The ANOVA of these *d'* values revealed a significant interaction of familiarity and condition, showing that blurred stimuli were better recognized when they were familiar. These results suggest that successful processing of configural information requires familiarity with the face whereas recognition of novel faces relies predominantly on the processing of featural information. In the course of familiarization the importance of configural information increases.

2.1. Introduction

Faces have attracted the attention of numerous scientists. They are a relevant class of stimuli and can be recognized with high accuracy. Long-term memory for faces seems to be very reliable; a face can be recognized with 90% accuracy after more than 50 years (Bahrick et al., 1975). However, relatively little is known about how we perceive novel faces. Imagine seeing a smiling face passing in a crowd. We only see the face for a very short moment, but often it can be remembered hours later. This memory may be based on representations other than those involved in the recognition of a familiar face.

Indeed, various studies provide evidence that familiar faces are processed differently than unfamiliar faces. For example, in a study by Burton, Wilson, Cowan, and Bruce (1999) personally familiar faces could be matched with high accuracy, even if the quality of the images was impoverished, whereas unfamiliar faces were matched poorly (see also Henderson, Bruce, & Burton, 2001). Burton et al (1999) showed their participants video clips of faculty members entering the psychology building, routinely collected by a security camera. The video clips were of rather poor quality, as is characteristic for security video systems. In the following test phase the participants were shown high-quality photographs of faces, half of which had been presented in the video clips. The task was to decide whether each of these photographs appeared in the videos. The participants performed significantly better if they were familiar with the target faces. The authors concluded that unfamiliar face recognition is mainly based on image specific details such as lighting condition and viewing angle. In contrast, recognition of familiar faces seems to be mediated by more generic representations which can be generalized over changes of image specific properties (Burton et al 1999). Buttle and Raymond (2003) showed that highly familiar faces are perceptually processed more efficiently than recently learned faces. Two pairs of faces were successively presented; one of the faces was changed in the second pair. The task was to detect the changed face; it did not involve explicit recognition or naming. Performance was significantly better if the changed face was famous. This finding is referred to as the superfamiliarity effect.

Several neuro-imaging studies also suggest different processing mechanisms of familiar and unfamiliar faces. In an fMRI study Rossion et al (2003) found activation differences in areas that are predominantly involved in early perceptual aspects of face processing and suggested that several basic functions of the face processing system, such as face detection, individual discrimination, and pre-semantic recognition, play a role in differentiating familiar faces from novel faces (Rossion, Schiltz, & Crommelinck, 2003; Rossion, Schiltz, Robaye, Pirenne, & Crommelinck, 2001). Other brain regions differentiating between familiar and unfamiliar faces include the pre-frontal lateral temporal, hippocampal, and parahippocampal regions bilaterally (Leveroni et al., 2000).

Findings from neuro-imaging and behavioural studies suggest a difference between processing of familiar and unfamiliar faces. But what is the nature of the mechanism that underlies the distinction between novel and familiar faces? It has been suggested that an important cause goes back to two types of information processing for featural and configural information (e.g., Diamond & Carey, 1986). Featural information is referred to as the information contained in the components of a face (i.e., eyes, nose, mouth etc); configural information is understood as the spatial relationship between the features. Diamond and Carey (1986) further differentiated configural information, distinguishing between first order and second order relational information. They described first order relational information as the basic arrangement of the parts, such as “the nose is above the mouth” and “the eyes are above the nose”, whereas second order relational information is defined by the metric distances between the parts. As all faces share first order relational information, high importance is ascribed to second order relational information. Some authors indeed claim that faces are processed exclusively on the basis of configural information (e.g., Tanaka & Farah, 1993; Farah et al., 1995) whereas others adopt a dual-code view, ascribing importance to both featural and configural information (Bartlett et al., 2003; Cabeza & Kato, 2000; Schwaninger et al., 2002).

Many authors made use of the so called face inversion effect (FIE) to impede configural processing. It has been shown that processing of second order relational information is hampered when the stimuli are presented upside-down (e.g., Leder et al., 2001; Maurer et al., 2002; Rock, 1973; Searcy & Bartlett, 1996; Farah et al., 1995; Yin, 1969). In his classic study, Yin (1969) found a much poorer performance for recognizing inverted faces as compared to other inverted objects, such as airplanes, houses, or stick figures of men in motion. Consequently, this phenomenon was termed the face inversion effect (FIE). The FIE is a robust phenomenon, it has been found when participants name, classify, or match photographs or drawings of faces (for an overview see Valentine, 1988). It was claimed that perception of faces relies predominantly on configural information; therefore the inversion effect is so pronounced with faces. In an fMRI study Leube et al. (2003) presented upright and inverted faces and the participants had to decide in which orientation the faces appeared. Brain activation was calculated for upright minus inverted faces and a signal change in the right superior temporal sulcus and right insula was found. The authors suggest that the right superior temporal sulcus and regions of the insular cortex may be associated with configural processing as no activation was found in these regions when faces were inverted.

The fact that featural and configural information are unequally sensitive to inversion can be interpreted in favour of two dissociable mechanisms processing featural and configural information. Indeed, recent findings support the dual-code view that featural and configural processing constitute two different routes to face recognition, but that configural coding is the hallmark of face processing (e.g., Cabeza & Kato, 2000; Schwaninger et al., 2002; Sergent, 1984; Tanaka & Sengco, 1997; see also Bartlett et al., 2003). Neuro-imaging studies revealed further evidence for the existence of two separate mechanisms. Rossion et al. (2000) found a hemispheric difference between a configural and featural processing strategy. When participants focused their attention on particular features of faces, they showed more activation in the left fusiform face area. When they relied more on configural mechanisms to process faces activation was larger in the right fusiform face area.

Various studies have suggested that configural information plays an increasing role with growing expertise with an object class (Diamond & Carey, 1986; Gauthier et al., 2000; Gauthier & Tarr, 1997; Gauthier et al., 1999; for a review see Rossion & Gauthier, 2002). Everyday we have to distinguish between faces of people we know and strangers. Thus, adults have had many years to develop expertise in face recognition (Mondloch, Geldart, Maurer, & Le Grand, 2003; Mondloch et al. (2002). But expertise has to be distinguished from familiarity. Our experience with faces has made us to experts in this object class, but we are only familiar with faces we know and have encountered many times. For personally familiar faces we have had sufficient time to carefully process the face and acquire representations which can be used to identify an observed face. For novel faces, however, there are neither featural nor configural representations from past encounters to rely on. In most previous studies investigating featural and configural information the faces were first presented in a learning phase, so that participants could build up representations of the faces. Sometimes faces of celebrities were used, of which the participants already had representations (e.g., Buttle & Raymond, 2003; Collishaw & Hole, 2002). If unfamiliar faces were used, the exposure time was at least 3 seconds (e.g., Cabeza & Kato, 2000; Collishaw & Hole, 2002; Leder & Bruce, 1998) or even longer (Haig, 1984; Leder & Bruce, 2000; Schwaninger et al., 2002; Tanaka & Farah, 1993). In the present study the possibility that the participants could acquire representations of the stimuli faces was minimized by only briefly presenting novel faces without any encoding phase prior to the study. The present study therefore differs from the studies mentioned above in that the stimuli were in fact novel; the exposure was limited to 1 second only.

The issue addressed in the present study was how faces are processed when seen for a brief duration only. I used scrambled and blurred stimuli to separately investigate the role of featural and configural information in novel faces. When a face is sufficiently blurred, the componential information of the features is reduced (e.g., the individual shape and texture of the eyes, mouth, nose). Likewise, configural information of a face is reduced, when its constituent parts are scrambled (Collishaw & Hole, 2000, 2002; Schwaninger et al., 2002; for a review see Rakover, 2002). Schwaninger et al. (2002) determined their blur level by simultaneously blurring and scrambling the faces and thus impairing configural and featural information. If scrambling and blurring faces are adequate manipulations to separately examine configural and featural processing, faces that are simultaneously blurred and scrambled should no longer be reliably recognized. This was indeed the case in the control experiment reported by Schwaninger et al. (2002). If faces are indeed special in a sense that they are predominantly coded configurally, then an advantage for a blurred face, as opposed to a scrambled face, is expected when briefly presented. Alternatively, if configural coding relies on the familiarity of a face, a briefly presented blurred face will not be reliably recognized.

All experiments reported here employ a same-different sequential matching task with blurred, scrambled and intact faces which could be either upright or inverted. In Experiment 1 two sequentially presented blurred or scrambled faces had to be matched. In Experiment 2 a blurred or a scrambled face had to be matched with an intact face. Experiment 3 tests the influence of familiarity on configural and featural face processing.

2.2. Experiment 1

Using scrambled and blurred faces as stimuli, the aim of Experiment 1 was to separately investigate the role of configural and featural information when processing novel faces. Four different conditions were tested. In the congruent conditions a blurred face had to be matched with another blurred face or a scrambled face had to be matched with another scrambled face. In the incongruent conditions a blurred face had to be matched with a scrambled face and vice versa.

If configural information plays a dominant role for processing novel faces, participants are expected to perform better when blurred faces have to be matched. If, on the other hand, configural information cannot be processed instantly, an advantage is expected for scrambled faces. Inversion is known to disrupt configural information to a much larger degree than featural information. Therefore I expected an inversion effect in the blurred condition but not in the scrambled condition.

2.2.1 Method

2.2.1.1 Participants

Sixteen participants (12 female / 4 male) ranging in age from 21 to 36 years volunteered to participate in Experiment 1. They were paid for their participation and all reported normal or corrected-to-normal vision.

2.2.1.2 Apparatus

The study was run on a 15.1” Pentium 4 portable Computer using Superlab Pro 2.0.2 running on Windows NT. The experiment took place in a dimly lit room. The participants were seated on a height-adjustable chair and responded by using a Cedrus© Response Pad (RB-520). They were asked to keep the head still on a head-rest, keeping a viewing distance of 500 mm constant. Each stimulus face appeared 95 mm wide and 125 mm high and thus subtended a visual angle of approximately 9.5° horizontally.

2.2.1.3 Stimuli

The stimuli were created from 60 photographs of Caucasian faces (26 male, 34 female) taken at the University of Zurich in the years 2000 and 2001. The models agreed on their photographs being used for experimental studies. The faces were of a neutral expression and were photographed from a frontal view. All faces were scaled to a standard size of 300 pixels across the width of the face at pupil level. For the blurred condition, the stimuli were created in three steps. First, colour information was discarded in the intact faces. This was done because colour does not contain any space related information. Second, the faces were blurred using a Gaussian filter with a sigma of 0.025 of image width in frequency space, using the equation exp(-f2 / (2*sigma2)). This blur level was slightly stronger than the blur level used by Schwaninger et al. (2002). I decided to increase the blur level to avoid ceiling effects. In a third step the faces were cut out with an elliptic tool provided by Adobe Photoshop 7.0 using soft contours (5 pixel feather). Thus the outer features of the faces such as head shape and hair line were discarded and all the faces appeared at the same size and shape (296 pixels wide, 385 pixels high). An example of a blurred stimulus is shown in Figure 1a.

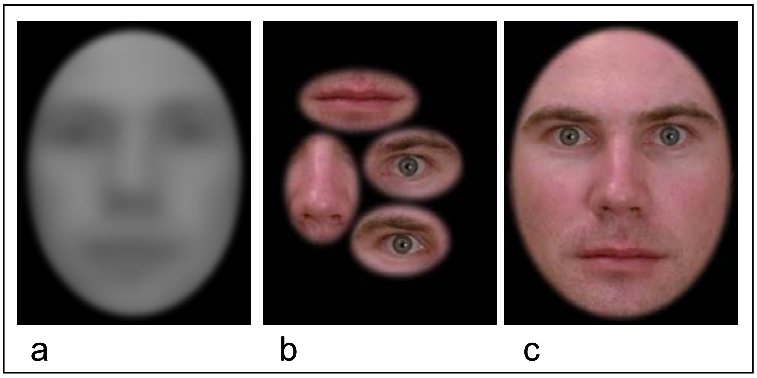

Figure 1. Examples of stimuli. a) blurred face used in Experiment 1 - 3; b) scrambled face used in Experiment 1 - 3; c) intact face used in Experiment 2 and 3.

Scrambled faces were created from intact faces in the following steps. Eyes, mouth and nose were cut out with the elliptic tool described above (eyes: 131 pixels wide and 95 pixels high, mouth 160 x 82, nose 98 x 145 pixels). These features were placed on a black background and scrambled in four different versions, to ensure that the features did not appear at the same location in each scrambled stimulus. Thus the location of each feature was not predictable. Each scrambled version was arranged so that no part was situated in its natural relation to its neighbouring part. The scrambled features were placed within the same area as the blurred stimuli, so they subtended to the same visual angle. An example of a scrambled stimulus can be seen in Figure 1b. All the stimuli could either appear upright or inverted.

2.2.1.4 Task and Procedure

Experiment 1 was a sequential matching task. A trial consisted of a fixation cross which appeared for 500 ms, followed by a stimulus face which was either scrambled or blurred and either upright or inverted. After 1000 ms the stimulus face disappeared and was replaced with

a random dot mask, to avoid afterimages of the stimuli. After 1000 ms the mask was replaced by the second face stimulus which was again either blurred or scrambled and either upright or inverted. After 1000 ms the second stimulus face disappeared and the screen turned blank until an answer key was pressed and the next trial began. Participants were told to respond as fast and as accurately as possible by pressing the appropriate key on the response pad. Thirty-two different trials were possible: a trial could be same or different, the first face could be blurred or scrambled and upright or inverted, and the second face could be blurred or scrambled and upright or inverted (2 x 2 x 2 x 2 x 2). When both faces were scrambled they never appeared in identical versions.
Prior to the experiment all participants gave informed consent. They received written and oral instructions and underwent 12 practice trials to ensure that they understood the task. None of the stimuli used in the practice trials occurred in the experiment proper. The experiment consisted of two blocks with 128 trials each encompassing all possible conditions. Different face pairs were used in each block and the order of the two blocks was counterbalanced across participants. Each block was approximately 10 minutes long. After the first block participants were able to take a break in order to regain their attention.

2.2.2 Results

The *d’* values were calculated for each participant in each condition by subtracting the z-transformed false alarm rates from the z-transformed hit rates. For false alarm rates of 0 and hit rates of 1 I used an approximation of the z-values -3 and 3, respectively. A total of 16 conditions resulted from combinations of scrambled (sc), blurred (bl), upright (upr) and inverted (inv) faces. For example, the condition in which an upright blurred face was followed by an inverted scrambled face will be abbreviated bl/upr-sc/inv. The average *d’* values are shown in Figure 2. One sample t-tests (two-tailed) on each condition revealed that all conditions differed significantly from 0, except in the conditions bl/inv-sc/upr ($d' = 0.104$, $t = 0.651$, $p = .525$) and bl/inv-sc/inv ($d' = 0.207$, $t = 0.879$, $p = .393$).

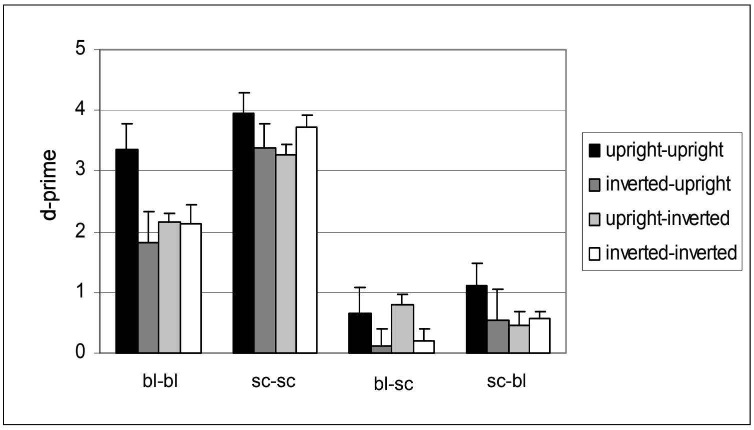

Figure 2. Mean d’ values for the 4 conditions in Experiment 1, the error bars depict standard errors.

Using a within-subjects design, a two-way ANOVA was carried out on the factors presentation mode (bl-bl, sc-sc, bl-sc, sc-bl) and orientation (upr-upr, inv-upr, upr-inv, inv-inv), revealing significant main effects of presentation mode, $F(3,45) = 45.71$, $MSE = 3.1$, $p < .001$, and orientation, $F(3,45) = 5.78$, $MSE = 1.32$, $p < .01$. The interaction presentation mode x orientation was not significant, $F(9,135) = 1.37$, $MSE = 1.04$, $p = .207$. Mean *d’* values for presentation mode were: bl-bl, M = 2.36, SE = 0.25; sc-sc, M = 3.58, SE = 0.36; bl-sc, M = .44, SE = 0.1; and sc-bl, M = 0.67, SE = 0.13. Post-hoc pair-wise comparisons (Bonferroni corrected) revealed a difference between the conditions with congruent presentation modes (bl-bl and sc-sc), SE = 0.3, p < .01, showing that two scrambled faces were better matched than two blurred faces. Moreover, both congruent conditions revealed significantly higher *d'* values than the non-congruent conditions, as revealed in post-hoc pair-wise comparisons (Bonferroni corrected) for bl-bl with bl-sc, SE = 0.2, p < .001, and for bl-bl with sc-bl, SE = 0.31, p < .001, for sc-sc with bl-sc, SE = 0.34, p < .001, and for sc-sc with sc-bl, SE = 0.44, p < .001. The incongruent conditions bl-sc and sc-bl did not differ, SE = 0.2, p = 1.000.

The different influence of orientation on scrambled and blurred faces was tested with 2 one-way ANOVAs computed separately for the two congruent conditions (bl-bl and sc-sc). A significant effect of orientation was found for blurred, $F(3,45) = 3.5$, $MSE = 2.05$, $p < .05$, but not for scrambled stimuli, $F(3,45) = 1.28$, $MSE = 1.22$, $p = .293$.

For the analysis of the reaction times (RTs) only the correct answers were included. Also, RTs over 3000 ms were discarded and treated as outliers. The RTs were subjected to a three-way ANOVA with the factors presentation mode (bl-bl, sc-sc, bl-sc, sc-bl), sameness (same, different), and orientation (upr-upr, inv-upr, upr-inv, inv-inv), which revealed a main effect of presentation mode, $F(3,45) = 16.31$, $MSE = 78336.26$, $p < .001$. The factors sameness and orientation did not reach significance; neither did any of the interactions. The mean RTs for presentation mode were as follows: bl-bl, M = 1339 ms, SE = 61.4, sc-sc, M = 1308 ms, SE = 69.6, bl-sc, M = 1498 ms, SE = 66.7, and sc-bl, M = 1493 ms, SE = 62.1. Post-hoc pair-wise comparisons (Bonferroni corrected) revealed that congruent matches were faster than incongruent matches, between bl-bl and bl-sc (SE = 29.668, p < .001) and between bl-bl and sc-bl (SE = 43.422, p < .05). Likewise, the comparisons of sc-sc with bl-sc (SE = 22.11, p < .001) and sc-sc with sc-bl (SE = 45.09, p < .01) reached statistical significance. The difference between bl-bl and sc-sc was not significant (SE = 28.47, p = 1.000). The mean RTs are depicted in Figure 3.

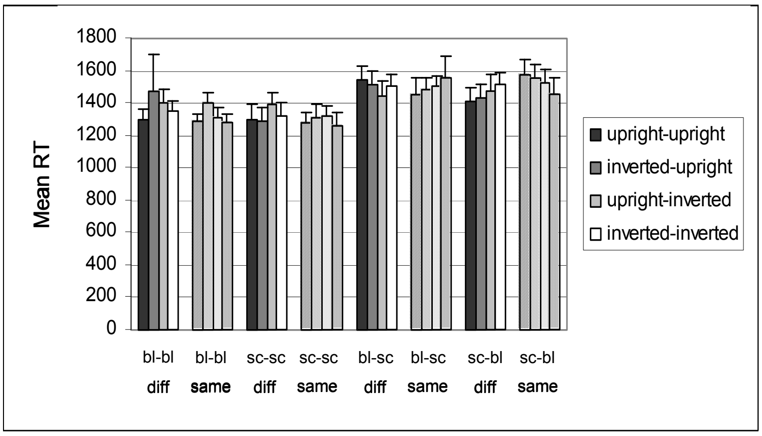

Figure 3. Mean RT values for the 4 conditions in Experiment 1, depicted separately for different and same trials. The error bars represent standard errors.

2.2.3 Discussion

Processing featural information evoked fewer errors, as can be concluded from the significantly higher *d'* values in the sc-sc condition compared to the bl-bl condition. This evidence goes in line with the hypothesis that configural information in novel faces cannot be processed as easily as featural information. When a face is presented for a brief duration only, the observer has to rely predominantly on featural information for later recognition.

The results show that both isolated configural and featural face information contained in the input stimulus can be processed reliably; the *d'* values of the bl-bl condition and the sc-sc condition were recognized above chance level. The fact that the RTs of the bl-bl and sc-sc conditions did not differ statistically suggests that there was no speed-accuracy trade off.

When blurred faces were involved, inversion hampered correct matching of two faces, whereas no inversion effect could be found for scrambled faces. This goes in line with several other studies, reporting that featural information is not as orientation sensitive as configural information (e.g., Leder et al., 2001; Maurer et al., 2002; Searcy & Bartlett, 1996; Farah et al., 1995; Yin, 1969). Furthermore, the effect of inversion for blurred faces provides evidence that no featural information was available to solve the task. The missing inversion effect for scrambled faces on the other hand suggests that configural information in these faces was in fact eliminated. Unexpectedly, all but two incongruent conditions were matched above chance level. This may suggest that scrambled and blurred faces do not contain mutually exclusive elements of the facial image. Alternatively, it could suggest that featural and configural processing is not entirely independent. Which of these explanations holds true can not be concluded on the basis of the present data. However, the fact that the congruent conditions were matched more accurately than the incongruent conditions denotes that featural and configural processing was substantially impaired through the manipulations.

Analysis of the RTs revealed that congruent matching (i.e., when the faces were manipulated by the same means) was faster than incongruent matching (scrambled faces with blurred faces and vice versa). This goes in line with the decreased *d'* values in the incongruent conditions. There was no effect of orientation, which suggests that for angles around 180° observers can flip the stimulus to the upright, instead of mentally rotating it in the picture plane (cf. Kanamori & Yagi, 2002; Murray, 1997).

Taken together, these findings suggest that featural information in novel faces is processed more efficiently than configural information. A possible explanation for this finding is that the formation of configural representations takes longer than forming featural representations. Therefore the recognition of novel faces relies predominantly on featural information.

2.3. Experiment 2

Experiment 1 showed that isolated configural or featural information can be processed. But, matching two stimuli containing either featural or configural information may rely on different strategies than actually using configural or featural information to recognize a new face which is intact. Furthermore, we mainly encounter intact faces in everyday situations. So the question addressed in Experiment 2 was whether natural novel faces can be recognized on the basis of isolated configural or featural information. The same design was used as in Experiment 1, except that the second face was always intact. An intact face contains both kinds of information, but not in isolated form. In the present task isolated configural and featural information have to be transferred to an intact face and any kind of direct matching can be ruled out. If novel face processing relies more on featural information, scrambled faces are expected to be matched more reliably to intact faces than blurred faces. Furthermore, no effect of inversion for scrambled faces was expected since featural representations are likely processed in an orientation-invariant manner.

2.3.1 Method

2.3.1.1 Participants

Fourteen participants (10 female / 4 male) ranging in age from 21 to 36 years volunteered to take part in Experiment 2. They were paid for their participation and all reported normal or corrected-to-normal vision.

2.3.1.2 Apparatus

The apparatus was the same as in Experiment 1.

2.3.1.3 Stimuli

The stimuli were the same as in Experiment 1, except that in addition to the scrambled and blurred stimuli un-manipulated, intact faces were used. The intact stimuli were cut out with the same elliptic tool as were the blurred stimuli, thus they were the same in size and shape. Figure 1c shows an example of an intact stimulus. Again, all faces were presented either upright or inverted.

2.3.1.4 Task and Procedure

The task and procedure was the same as in Experiment 1, except that the second stimulus face was always an intact face. Sixteen different trials were possible (2 x 2 x 2 x 2). The experiment encompassed a total of 256 trials.

2.3.2 Results

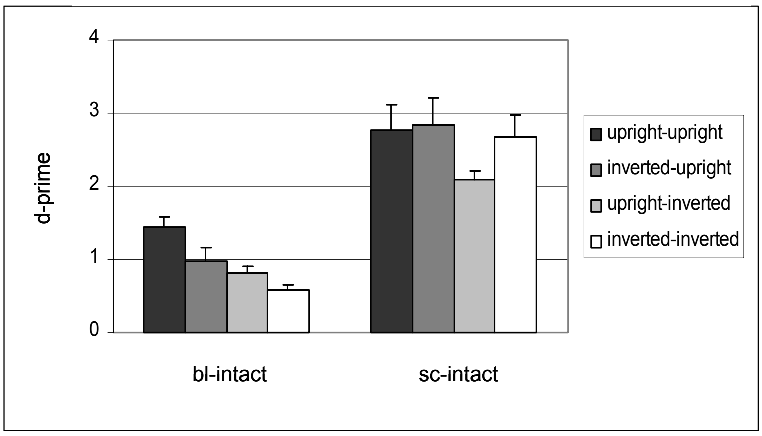

Figure 4. Mean d' values for the 2 conditions in Experiment 2, the error bars depict standard errors.

As in Experiment 1 the *d'* values were calculated for each participant in each condition. The mean *d'* values are depicted in Figure 4. One sample t-tests (two-tailed) on each condition revealed that all conditions differed significantly from 0.

Using a within-subjects design, the *d'* values were subjected to a two-way ANOVA with the factors presentation mode (bl, sc) and orientation (upr-upr, inv-upr, upr-inv, and inv-inv), revealing a significant main effect of

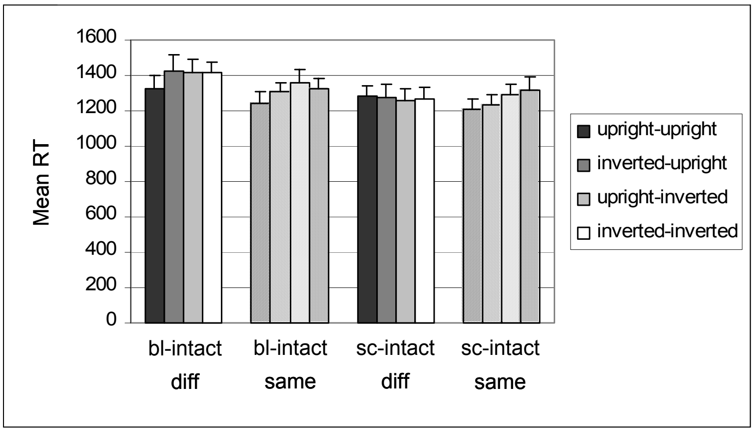

Figure 5. Mean RT values for the 2 conditions in Experiment 2, depicted separately for different and same trials. The error bars represent standard errors.

presentation mode, $F(1,13) = 73.32$, $MSE = 1.03$, $p < .001$, and orientation, $F(3,39) = 3.51$, $MSE = 0.69$, $p < .05$. A post-hoc pair-wise comparison (Bonferroni corrected) of the presentation mode was significant, SE = 0.19, p < .001, showing that scrambled faces (mean *d'* = 2.6, SE = 0.19) were better matched with intact faces than blurred faces (mean *d'* = 0.96, SE = 0.06). The interaction presentation mode * orientation was not significant, $F(3,39) = 2.24$, $MSE = 0.51$, $p = .099$.

To separately investigate the effect of orientation on scrambled and blurred faces, two one-way ANOVAs were computed for the blurred conditions and the scrambled conditions, revealing a significant effect of orientation for blurred, $F(3,39) = 8.4$, $MSE = 0.23$, $p < .001$, but not for scrambled stimuli, $F(3,39) = 1.68$, $MSE = 0.98$, $p = .186$.

As in Experiment 1 only RTs under 3000 ms of the correct answers were taken into account. A 2 x 2 x 4 ANOVA on the RTs with the factors presentation mode (bl, sc), sameness (same, different), and orientation (upr-upr, inv-upr, upr-inv, inv-inv) revealed a main effect of presentation mode, $F(1,13) = 11.66$, $MSE = 35107.21$, $p < .01$, and sameness, $F(1,13) = 4.83$, $MSE = 27299.6$, $p < .05$. It took the participants longer to respond when the intact face was preceded by a blurred face (M = 1352 ms) compared to when the intact face was preceded by a scrambled face (M = 1266 ms). The RTs were shorter when the two faces were the same (M = 1285 ms) than when they were different (M = 1333 ms). There was no effect of orientation; neither were there any interactions. The average RTs are depicted in Figure 5.

2.3.3 Discussion

The *d'* values were significantly larger in the scrambled condition than in the blurred condition, suggesting an advantage for featural information in the processing of novel faces. The advantage for featural information is also reflected in the shorter RTs in the scrambled condition. This finding is not in line with several studies providing evidence that face processing relies mainly on configural information (Diamond & Carey, 1986; Farah et al., 1998; Mondloch et al., 2002; Tanaka & Farah, 1991, 1993; Farah et al., 1995). It could be argued that in the scrambled condition a picture-matching strategy could be adopted which accounts for the higher *d'* values in this condition, because the features are contained unchanged in the intact faces. However, it has to be noted that the scrambled and intact faces are not presented simultaneously. Therefore a pure perceptual matching strategy could not have been adopted. Moreover, also the configural information contained in the blurred faces remains unchanged in the intact faces. Therefore, a similar kind of matching strategy could have been used just as well in the blurred conditions. In one condition parts had to be matched with a whole face, in the other condition the configuration had to be matched with the spatial relations within the whole face.

As expected, and in line with the findings of Experiment 1, inversion disrupted recognition in the blurred condition, but had no effect in the scrambled condition. A missing inversion effect in the scrambled condition indicates that no configural information could be used to recognize the face. Above all, this finding suggests that featural representations are orientation-invariant; the individual features could be processed independent of orientation. The large inversion effect in the blurred condition, however, suggests that the information available in blurred faces was predominantly configural.

The analysis of the RTs supports the hypothesis that featural information is processed more easily in novel faces. Scrambled faces were matched with intact faces significantly faster than blurred faces. As in Experiment 1 no effect of inversion was found for the RTs.

2.4. Experiment 3

The results of Experiment 2 suggest that novel faces are processed primarily on featural information contained in the facial parts. Experiment 3 was concerned with the question whether configural information becomes more important when a face is familiar, as suggested for example by Buttle and Raymond (2003). Two sets of faces were used, one set was extensively studied before the experiment, the other set consisted of novel faces. If configural information becomes more important during the course of familiarization, an increase of *d’* values would be expected in the blurred condition for learned faces, whereas the *d’* values in the scrambled condition were expected to be comparable for both learned and novel faces.

2.4.1 Method

2.4.1.1 Participants

Eighteen participants (9 female / 9 male) ranging in age from 22 to 33 years (mean 28.9) took part in Experiment 3. They were paid for participation and all reported normal or corrected-to-normal vision.

2.4.1.2 Apparatus

The apparatus was the same as in Experiment 1 and 2.

2.4.1.3 Stimuli

The stimuli were the same as in Experiment 2. Figure 1 shows examples of the stimuli. In Experiment 3 all stimuli were presented in upright orientation only.

2.4.1.4 Task and Procedure

The task was the same as in Experiment 2. Twenty faces were randomly chosen for the familiar set and were included in a learning session prior to the experiment proper. In the learning session, participants were subsequently presented 20 intact faces. After 10 seconds the face disappeared and the next face appeared as soon as the centre button was pressed. The learning session was repeated three times with a different presentation order of the faces, which was randomized online. After the three learning sessions an old-new recognition test was interposed. In this test the 20 learned faces were subsequently presented together with 20 distractor faces. For each face participants were required to decide whether the face was "old" (i.e., whether it was one of the faces from the learning session) or whether it was "new". In the experiment proper 40 pairs (20 same, 20 different) were created from the 20 old faces and 40 pairs (20 same, 20 different) were created from 20 new faces. These new faces were other faces than the distractor faces used in the learning test. Thus, the experiment consisted of 80 face pairs.

2.4.2 Results

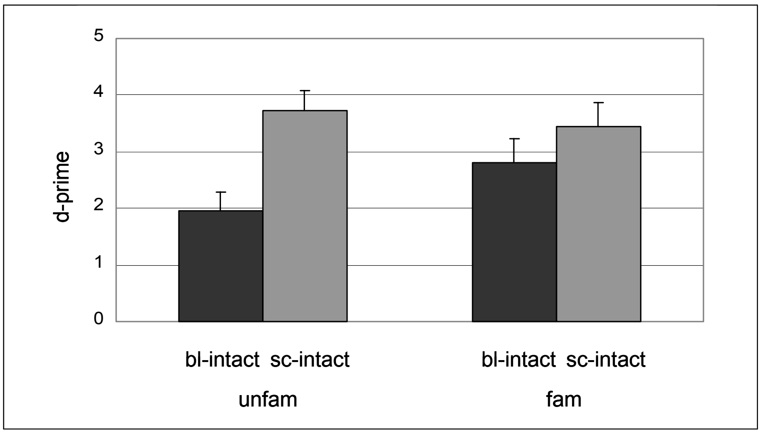

Figure 6. Mean d' values for the 2 conditions in Experiment 3, depicted separately for familiar and unfamiliar trials. The error bars depict standard errors.

Participants were excluded from the analyses if they failed to recognize at least 70% of the studied faces in the old-new recognition task. By ensuring that more than 70% of the faces were recognized accurately, it could be assumed that these faces were sufficiently learned. Thus, two participants were discarded from the analyses. Only the data of the remaining 16 participants (9 female, 7 male) were included in the following analyses (mean recognition rate 90.3%). As in the previous experiments *d'* and RTs were analyzed. The mean *d'* values are depicted in Figure 6. One sample t-tests (two-tailed) on each condition revealed that all conditions differed significantly from 0 (all $t > 5.6$, $p < .001$). Using a within-subjects design, the *d'* values were subjected to a 2 x 2 repeated measures ANOVA with the factors presentation mode (bl, sc) and familiarity (old, new), revealing a main effect of presentation mode, $F(1,15) = 8.008$, $MSE = 2.93$, $p < .05$. There was no effect of familiarity, but the two-way interaction of presentation mode and familiarity reached statistical significance, $F(1,15) = 5.989$, $MSE = .886$, $p < .05$. This interaction was due to the blurred condition, as two sample t-tests (two-sided) between the *d'* values of familiar and unfamiliar faces were significant only in the blurred condition, $t(15) = 2.597$, $p < .05$. In the scrambled condition this comparison was not significant. Furthermore, the comparison of the blurred and the scrambled condition for novel faces was significant (T = 4.332, p = .001), for familiar faces the same comparison was not significant (T = 1.145, p = .27).

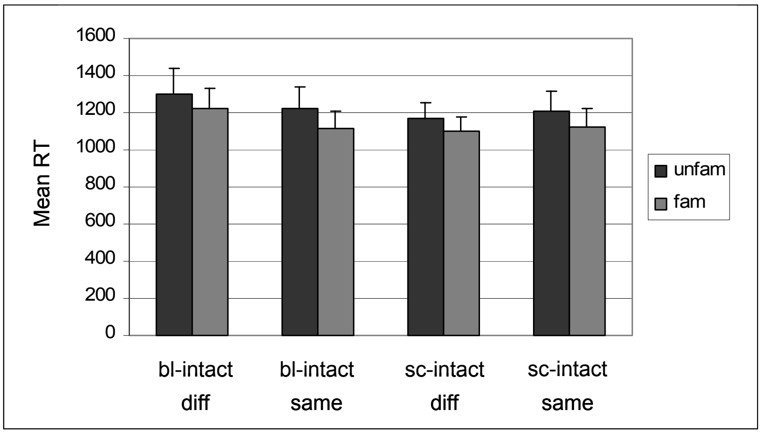

Figure 7. Mean RT values for the 2 conditions in Experiment 3, depicted separately for familiar and unfamiliar trials. The error bars represent standard errors.

As in the previous experiments only RTs under 3000 ms of the correct answers were taken into account. A 2 x 2 x 2 ANOVA on the RTs with the factors presentation mode (bl, sc), familiarity (familiar, unfamiliar), and sameness (same, different) revealed a main effect of familiarity, $F(1,15) = 14.48$, $MSE = 15901.745$, $p < .01$. It took the participants longer to respond when the faces were unfamiliar (M = 1225 ms) compared to when the faces were familiar (M = 1140 ms). The main effect of presentation mode did not reach significance; neither did any of the interactions. The average RTs are depicted in Figure 7.

2.4.3 Discussion

The most important finding of Experiment 3 was that familiarity selectively improved recognition of blurred faces. For recognizing scrambled faces learning a face beforehand had no effect. This finding strongly supports my hypothesis that for processing novel faces featural information plays a dominant role and confirms previous findings of Buttle and Raymond (2003) who claimed that configural information is more important in familiar faces than in unfamiliar faces.

When a face becomes familiar, the configuration becomes more and more important. For unfamiliar blurred faces *d'* values were comparable to the upr-upr condition in Experiment 2. For unfamiliar scrambled faces the *d'* values in Experiment 3 were somewhat higher than in Experiment 2. This could be due to the fact that in Experiment 3 the unfamiliar pairs were created from only 20 faces. Blurred and scrambled conditions did not differ when the faces were familiar, but differed significantly for novel faces. It becomes clear that configuration is an important source of information to recognize both novel and familiar faces. However, the relative importance of configural information is reduced in novel faces. The RTs were significantly shorter for familiar faces. This was the case for scrambled and blurred conditions. It is likely that existent representations of features and configural information speed up the process of face recognition.

2.5 General Discussion

The main finding of this study is that featural information plays a larger role than configural information in recognizing novel faces. A significant advantage was found for scrambled faces, which contained isolated featural information. Blurred faces, which contain isolated configural information, were recognized less accurately when presented briefly. For familiar faces (i.e., for faces that were extensively learned before the experiment) configural information showed growing importance. The present results suggest that in order to effectively process configural information, a face has to be at least somewhat familiar. In other studies, faces had to be learned during a certain amount of time (e.g. Cabeza & Kato, 2000; Leder & Bruce, 1998, 2000; Schwaninger et al., 2002; Tanaka & Farah, 1993; Farah et al., 1995) or highly familiar famous faces were used (e.g., Buttle & Raymond, 2003; Collishaw & Hole, 2000, 2002). In both cases it was ensured that the observers had configural representations of the faces before the experiment began. In Experiments 1 and 2 of the present study novel faces were presented for one second only. In Experiment 3 I specifically tested the influence of familiarity on the processing of configural information. The results clearly support the view, that the importance of configural information increases in familiar faces, whereas featural information is processed equally well in novel and familiar faces. Although configural information could be processed in novel faces (blurred faces were matched above chance level), the relative importance of the configuration was reduced compared to familiar faces. On the basis of the present study I come to slightly different conclusions than a similar study by Collishaw and Hole (2000). They also used scrambled and blurred faces to separately analyse featural and configural processing and reported no difference between celebrities and unfamiliar faces. Moreover, they found no difference between the scrambled and blurred conditions. However, the participants in that study saw the unfamiliar faces once before during 3 seconds. This brief exposure might have sufficed to build a configural representation. Collishaw and Hole's unfamiliar condition is therefore comparable to the familiar condition in Experiment 3, where no difference between blurred and scrambled faces was found. Another reason why Collishaw and Hole (2000) found relatively high *d'* values for blurred faces could lie in the fact that they used faces where external features such as head shape and hairline were visible. Such features may facilitate recognition, as they are very characteristic for an individual face. I discarded such information by presenting each face in the same oval shape, thus the participants only could rely on the spatial relationship between facial parts.

The fact that novel and familiar faces are processed differently is consistent with the assumption of separate processing mechanisms for featural and configural information. For a novel face, no configural representation exists. Instead, we rely on featural information, which can be processed easily with only limited experience with the face. When a face becomes more familiar the configuration of that face is learned and a configural representation is formed. For novel faces the processing of featural information dominates and once configural representations have been formed the efficiency of configural processing increases.

Various authors have suggested that expertise refines configural processing (Diamond & Carey, 1986; Gauthier et al., 2000; Gauthier & Tarr, 1997; Gauthier et al., 1999; Rossion & Gauthier, 2002). On the basis of the present data I claim that unfamiliar, novel faces primarily activate a featural mode of analysis. This is in line with the findings of Buttle and Raymond

(2003) who report that it is growing familiarity with a face which is crucial for configural processing.

Blurred faces showed an effect of inversion, whereas scrambled faces did not. This finding indicates that configural information was destroyed in the scrambled stimuli and preserved in the blurred stimuli, as configural information in faces is responsible for the FIE. More important, a missing FIE for scrambled faces suggests that featural representations are orientation-invariant; features are not only processed fast, but also regardless of their orientation. Furthermore, the fact that inversion affects scrambled and blurred stimuli to a different degree provides further support for two dissociable processing mechanisms for configural and featural information. Whether these two mechanisms can be dissociated anatomically is not yet fully clear. Findings of Rossion et al (2000) show an asymmetric hemispheric activation, with a higher activation in the right hemisphere for configural processing and in left hemisphere for featural processing. On the basis of a distributed neural network (Haxby et al., 2001; Haxby et al., 2000; Ishai, Ungerleider, & Haxby, 2000; Ishai et al., 1999) no dissociable mechanisms are expected. Instead, the difference between configural and featural processing is suggested to be the result of quantitative activation differences within the same network. It has to be the issue of future studies to further resolve whether featural and configural processing follow two anatomically distinct pathways.

A difference between familiar and unfamiliar faces is that familiar faces have been visually explored extensively and therefore well practiced scan paths of eye movements have been acquired. These scan paths play an active role in visual recognition (Liversedge & Findlay, 2000); yet this information is available only when sequences of eye movements have been memorized before. In this study novel faces were presented so briefly that participants had no scan path to be recalled from long-term memory. It is possible that eye scan paths reflect the expertise specific to the processing of configural information. Indeed, neuro-imaging has revealed that the neuronal resources used for eye movements changes with the familiarity of the scan path to be executed (Grosbras et al., 2001).

The underlying mechanisms for processing novel faces are also interesting for forensic concerns, such as eye-witnessing. Situations are imaginable where a face is only seen for an instant, but can still be remembered clearly. For example, some witnesses of hit-and-run accidents report that they explicitly remember the face of the absconding driver, although it was only seen for a trickle of a moment. The present data suggest that such memories are based on featural representations. This finding may also have important implications for the use of security surveillance systems. Video images captured by security cameras are generally of low resolution. Only familiar people captured by the camera are recognized with high accuracy, unfamiliar people are identified very poorly (Bruce, Henderson, Newman, & Burton, 2001; Burton et al., 1999). The usage of high resolution cameras would enable the security surveillance to process more featural details, as offenders are unlikely familiar to the security guards.

Nevertheless, this proposition is not completely consistent with findings of studies investigating memory of faces. Schooler and Engstler-Schooler (1990) reported that participants demonstrated an impaired ability to recognize a person if they described the person verbally prior to the recognition task. This phenomenon has been termed the verbal-overshadowing effect. Macrae and Lewis (2002) interpreted the verbal-overshadowing effect on the basis of featural and configural processing strategies. They suggest that memory of a face is disrupted when a featural processing strategy is triggered. In contrast, adopting a configural processing strategy enhances the accuracy of face recognition. At first sight my results do not seem in line with studies on the verbal-overshadowing effect. It has to be noted, however, that in studies on the verbal-overshadowing effect participants saw 30 seconds of a videotaped simulated bank robbery, which was probably long enough to form a configural representation of the robber. In this study the encoding time was 1 second and it seems that this was too short to efficiently process configural information. Instead, memory had to rely exclusively on featural representations.

Taken together, the results of the present experiments propose a dual-code view that both configural and featural processing contribute to face recognition. For novel faces configural processing is not as powerful as featural processing, therefore featural information plays a predominant role when processing novel faces.

3. Face Imagery Is Based on Featural Representations

Abstract

The role of featural and configural representations in face imagery and face perception was investigated using blurred and scrambled faces. By means of blurring, featural information is hampered; by scrambling a face into its constituent parts configural information is lost. Prior to the experiment ten faces (5 female, 5 male) were learned together with the sound of a name. In the imagery condition a name was presented and participants were required to imagine the corresponding face as clearly and vividly as possible. Then, either a blurred or scrambled face was displayed. In the perception condition a scrambled or blurred face and a name were presented simultaneously, thus no facilitation via mental imagery was possible. In both conditions participants had to decide whether the name belonged to the face or not. Analyses of the hit values showed that in the imagery condition scrambled faces were recognized significantly better than blurred faces whereas there was no such effect for the perception condition. The results suggest that face imagery is based predominantly on featural representations whereas perceiving a learned face relies on both featural and configural representations.

3.1 Introduction

Imagine your history teacher back in your old schooldays. It may be a long time, but still the teacher's face can be imagined quite vividly. The pointed nose, the bushy eyebrows behind those shell-rimmed glasses, the thin hair are unforgettable. Needless to say that it feels different when we actually look at a photograph of this teacher as it may be shown around during the next class reunion. But what exactly is the difference? A striking distinction concerns the source that triggers a percept or a mental image. A percept has its origin in the stimulus whereas a mental image is evoked internally, based on previously memorized information. Therefore, it has often been suggested that images never give an impression of novelty, because we already know what we imagine. Other in perception: perception can teach me new things. In his work on the imaginary, Jean-Paul Sartre puts it this way: "if I give myself in image the page of a book, I am in the attitude of the reader, I look at the printed lines. But I do not read. And, at the bottom, I am not even looking, because I already know what is written" (Sartre, 1940, trans. 2004, p 10). In the context of face recognition this means that when I imagine my former teacher I will not have to undergo any recognition. The identity of the face is already known prior to the generation of the image. In contrast, when I see him on the street I will have to decide whether it is really him or not.

Despite these apparent differences various neuro-imaging studies on mental imagery of faces suggest that visual imagery evokes – at least partly - similar activation as when the faces are in fact perceived (Farah et al., 1988; Ishai et al., 2002; Ishai, Ungerleider, & Haxby, 2000). In an fMRI study, Ishai and colleagues (Ishai, Ungerleider, & Haxby, 2000) found content-related activation in extra-striate cortex and ventral temporal cortex when the participants visually imagined faces, houses, and chairs. It is noteworthy, however, that Ishai et al. (2000) also found some activity restricted to visual imagery in parietal and frontal cortex (see also Mechelli, Price, Friston, & Ishai, 2004).

Further evidence that imagery and perception of faces underlie similar neural mechanisms comes from case studies with prosopagnosic patients, which revealed that an impairment of face recognition is often accompanied with the disability to mentally visualize faces (Charcot & Bernard, 1883; Young et al., 1994; Young & Van De Wal, 1996). Some reports, however, have described prosopagnosic patients with intact face imagery (e.g., Bodamer, 1947; Pallis, 1955).

To further elicit the relation between imagery and perception of faces, Cabeza and colleagues (Cabeza, Burton, Kelly, & Akamatsu, 1997) conducted a priming study with healthy participants. They found that imagined faces prime imagined faces and seen faces prime seen faces, but they found no priming between seen and imagined faces. This led the authors to the conclusion that imagery is not merely a weak form of perception, thus favouring a view that imagery and perception rely on partly distinct processes. However, their perception and imagery conditions were not directly comparable. While they used a familiarity judgement as perception task, a speeded imagery test was used for the imagery task in which participants had to make judgements about the appearance of celebrities. The missing priming effect between seen and imagined faces may therefore be a result of task inconsistency. Moreover, Cabeza et al. (1997) only analysed response latencies because their design did not allow for any statement concerning accuracy. I will come back to this issue in the discussion section.

Taken together, a wealth of knowledge suggests that face imagery and face perception involve partly the same neural mechanisms (e.g., Farah et al., 1988; Ishai et al., 2002; Ishai, Ungerleider, & Haxby, 2000). It has to be noted, however, that the number of studies on face perception and face imagery is not balanced; far more studies have investigated face perception. Many of these studies differentiate between processing of configural and featural face information (e.g., Bartlett et al., 2003; Cabeza & Kato, 2000; Farah et al., 1995; Farah et al., 1998; Schwaninger et al., 2002; Tanaka & Farah, 1993; see also chapter 2). Featural information is referred to the constituent elements of a face (i.e., eyes, nose, mouth) whereas configural information is understood as the spatial relationship between these parts. Many authors have provided evidence that featural and configural information can be activated independently to recognise faces (e.g., Bartlett et al., 2003; Schwaninger et al., 2002) and it has been suggested that configural information plays a dominant role in face perception (e.g., Cabeza & Kato, 2000; Diamond & Carey, 1986; Farah et al., 1995; Farah et al., 1998; Schwaninger et al., 2002; Tanaka & Farah, 1993).

Assuming that perception and mental imagery indeed share some common mechanisms, I pursued the aim to investigate whether in mental imagery of faces configural and featural information can be similarly dissociated. Indeed, Ishai and colleagues (Ishai et al., 2002) found differential activation when participants attended to the features or the whole of imagined faces. Specifically, they found increased activation in the right intraparietal sulcus (IPS) and inferior frontal gyrus (IFG) when participants focussed on the features of the imagined face. This finding suggests that a dissociated neural mechanism processes featural information. While in perception configural information seems to play a predominant role, at least for familiar face recognition (e.g., Cabeza & Kato, 2000; Diamond & Carey, 1986; Farah et al., 1995; Farah et al., 1998; Schwaninger et al., 2002; Tanaka & Farah, 1993), this might not be the case in mental imagery. People asked to imagine a familiar face most likely describe the face by the features and not by configural characteristics. Much more likely they would mention the bushy eyebrows and the thin hair of the history teacher rather than the configural characteristics, such as his inter-eye distance is ¾ the distance between his mouth and eyes. However, when asked to verbally describe a mental image, people are likely to characterize a visual mental image of a face as fuzzy or blurred, suggesting that people may not be able to activate in imagery precise representations of facial parts after all.

The purpose of the present study was to investigate featural and configural representations in mental imagery and compare them to the role they play in perception. The importance of configural and featural representations in mental imagery and perception was ascertained by testing face recognition by means of scrambled and blurred stimuli. By scrambling the constituent parts of a face, global configural information contained in the face is destroyed. By blurring a face the detail featural information contained in the parts is hampered. These manipulations enable independent investigation of featural and configural information. I ascertained whether a mental image of a face primes featural or configural information, or both.

3.2 Experiment 4

3.2.1 Method

3.2.1.1 Participants

Twenty-four healthy participants (12 male / 12 female) ranging in age between 19 and 33 years (mean 25 years) took part in this experiment. Four participants reported to be left handed and all had normal or corrected to normal vision. All participants gave informed consent and were either paid for their participation or received course credits. The participants were treated according to the Declaration of Helsinki (1991).

3.2.1.2 Apparatus

The study was run on a 15.1” Pentium 4 portable Computer using Superlab Pro 2.0.2 running on Windows NT. The experiment took place in a quiet, dimly lit room. The participants were seated in a height-adjustable chair at a distance of 500 mm which was maintained by a headrest. They responded by using a Cedrus© Response Pad (RB-520). Each stimulus face appeared 95 mm wide and 125 mm high and thus subtended a visual angle of approximately 9.5° horizontally.

3.2.1.3 Stimuli

The stimuli were created from photographs of faces taken at the University of Zurich. The photographs were taken frontally and the faces were of a neutral expression. All faces were scaled to a standard size of 300 pixels across the width of the face at pupil level. The intact stimuli were cut out with an elliptic tool provided by Adobe Photoshop 7.0 using soft contours (5 pixel feather). Thus the outer features of the faces such as head shape and hair line were discarded and all the faces appeared at the same size and shape (296 pixels x 385 pixels). The target stimuli were given five letter names (e.g., Peter), which were presented acoustically and visually during the study phase. The names were typed in bold letters below the face. Figure 1 shows an example of an intact stimulus.

The blurred stimuli were created in two steps. First, colour information was discarded in the photographs. In a second step the faces were blurred using a Gaussian filter with a sigma of 0.025 of image width in frequency space, using the following equation exp(-f2 / (2*sigma2)). Using the same elliptic tool as for the intact stimuli the outer features were discarded. Thus the blurred stimuli were the same size and shape as the intact stimuli. An example of a blurred stimulus is shown in Figure 1.

Scrambled stimuli were created from the intact faces in following steps. Eyes, mouth and nose were cut out with the elliptic tool described above (eyes: 131 pixels x 95 pixels, mouth 160 x 82, nose 98 x 145 pixels). These features were placed on a grey background and scrambled in four different versions. Each version was arranged so that no part was situated in its natural relation to its neighbouring part. The scrambled features were placed within the same area as the intact and blurred stimuli, so they subtended to the same visual angle. An example of a scrambled stimulus can be seen in Figure 1.

3.2.1.4 Task and Procedure

The participants were given written and oral instructions. Prior to the experiment they underwent a demonstration version of the experiment, which consisted of shortened versions of the blocks described below. None of the stimuli used in the demonstration trials appeared in the experiment proper. The experiment started with a learning block. Participants learned the names of ten target faces. Each face was successively presented together with a name (e.g. “Peter”). The name was presented acoustically via headphones and visually in bold letters below the face. Participants were told to precisely memorize the face with its name so that they can later form a mental image that matches the original as precisely as possible. Half of the target faces were female, the other half male. The face

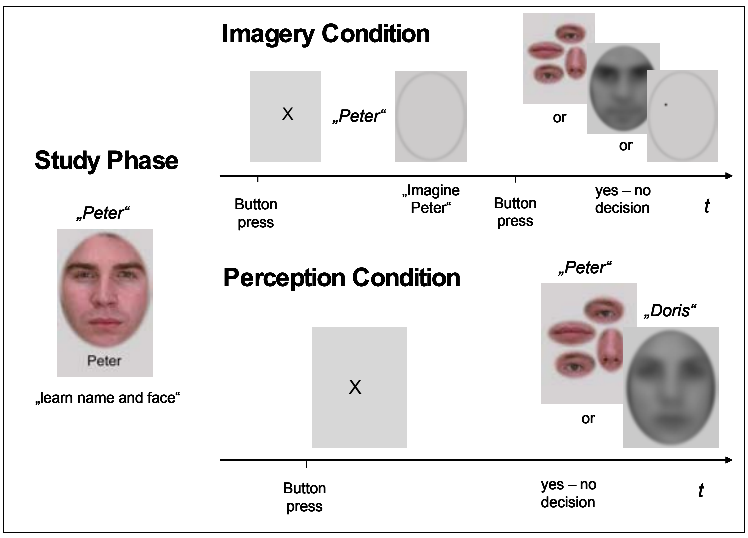

Figure 8: Examples of trials in the study phase, imagery and perception conditions. Italic words were presented acoustically.

was visible until a button was pressed. Then the screen went blank during which the participants were told to hold on to the image. As soon as the mental image started to fade, participants pressed the button again and the face reappeared and participants could correct and consolidate their image. On another button press the face disappeared anew and a fixation cross appearing for two seconds signalized the appearance of the next target face. A minimum of two study phases were carried out. To make sure that the participants learned the faces sufficiently, a naming task was carried out after the study phase. All target faces were presented subsequently and the participants had to name each face. If participants did not name all faces correctly a further study session was accomplished, until all ten faces were named correctly. To further practice mental imagery of faces another training block was included (imagery practice block). In this block each trial started with a fixation cross appearing for 1 second. Then a name was presented via headphones together with an oval shape indicating the array in which the face was to be imagined. Participants were requested to visualize the appropriate face as vividly as possible and fit the image onto the grey array. When the image was generated participants pressed a button, what made a small dot appear within the oval shape. This dot was either at the exact location where eye, nose or mouth would appear or 1 cm lower or higher than the feature. The participants then had to decide whether or not this dot would appear on a facial feature (eyes, nose, mouth) of the imagined face. As the location of the dot was determined individually for each face, the task required a highly accurate and vivid visual image of each face. After each answer the appropriate face appeared together with the dot, thus the participants were given direct feedback on their answers. This feedback enabled the participants to correct their mental image.

The experiment proper consisted of an imagery block (imagery condition) and a perception block (perception condition). The order of these blocks was counterbalanced across participants. The imagery condition was comparable to the imagery practice block, except that instead of the red dot either a blurred or a scrambled face appeared for one second. In a yes/no decision task participants had to decide whether the scrambled or blurred face corresponded to the face they imagined or not. In half of the trials the face and the name corresponded, in the other half name and face did not correspond. To control whether participants really mentally visualized the faces, ten tasks of the imagery practice block were included at random intervals, and participants had to decide whether or not the dot would appear on the location of the left or right eye, nose, or mouth, were it visually presented.

In the perception condition a trial started with a fixation cross followed by a blurred or scrambled face. At the same time, a name was presented via headphones. The task was to decide as quickly and as accurately as possible whether the presented name belonged to the blurred or scrambled face. As in the imagery condition half of the trials were same and half were different. The experimental design is shown in Figure 1. After the experiment participants completed the Vividness of Visual Imagery Questionnaire (VVIQ, Marks, 1973).

3.2.2 Results

Accuracy: The number of correctly matched faces was analysed. In the imagery condition the mean hit rate was 88.3% for scrambled faces and 67.9% for blurred faces. In the perception condition the mean hit rate was 75.0% for scrambled faces and 69.6% for blurred faces. The mean hit rates are depicted in Figure 2. A 2 (imagery, perception) x 2 (scrambled, blurred) analysis of variance (ANOVA) of the hit rates revealed a significant effect of information (scrambled, blurred), $F(1,23) = 19.005$, $MSE = 0.021$, $p < .001$. There was no main effect of condition, but condition and information interacted, $F(1,23) = 8.171$, $MSE = 0.017$, $p < .01$. Post-hoc *t*-tests (two-tailed) revealed that scrambled trials did not differ from blurred trials in the perception condition, but differed significantly in the imagery condition, $t = 5.086$, $p < .001$. Furthermore, the hit rates of blurred trials did not differ in the two conditions, but scrambled trials showed significantly higher hit rates in the imagery condition than in the perception condition, $t = 2.693$, $p < .05$.

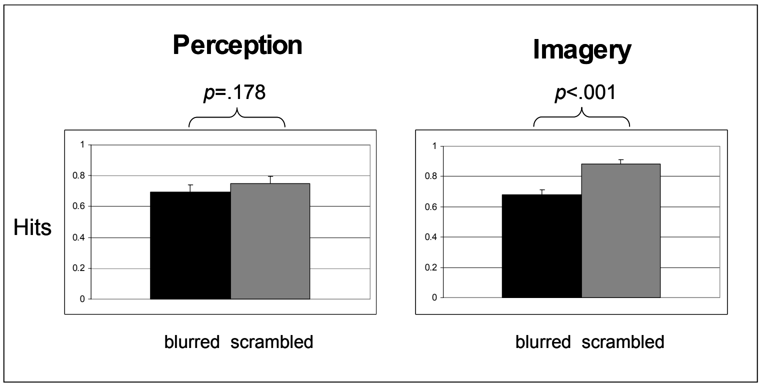

Figure 9: Mean d' values for scrambled and blurred trials. Left panel = perception condition, right panel = imagery condition. Error bars depict standard errors of the mean (SEM).

Reaction Times: Reaction times (RTs) that were 3000 ms or longer were treated as outliers and were not included in the analyses. Thus, less than 1.8 % of the trials in the imagery condition were excluded and less than 1.9% of the trials were excluded in the perception condition. Also, only RTs of correct answers were considered. In the imagery condition the mean RT was 1762 ms for blurred faces and 1721 ms for scrambled faces. In the perception condition the mean RT was 1739 ms for blurred faces and 1766 ms for scrambled faces. A 2 x 2 ANOVA of the RTs revealed no significant effects, neither for condition (imagery, perception), $F(1,23) = 0.024$, $MSE = 120362$, $p = .878$, nor for information (scrambled, blurred), $F(1,23) = 0.03$, $MSE = 37546$, $p = .865$. There was no interaction of condition x information. This finding suggests that there was no speed accuracy trade-off.

Control Condition: To ascertain whether participants were able to form a mental image of the test faces the *d'* values of the control condition were calculated by subtracting the z-transformed false alarm rates from the z-transformed hit rates. The mean *d'* value in the control condition of the experiment was 0.84 (SEM = .35). A one-sample t-test revealed that the *d'* values differed significantly from 0, T (23) = 2.413, p < .05, indicating that participants performed above chance level. The control condition was designed specifically to assess participants' performance in accurately visualizing the faces.

Questionnaires: The mean VVIQ scores ranged from 1.56 (clear and reasonably vivid image) to 3.31 (moderately clear and vivid image). Rating 1 indicated an image that is 'perfectly clear and as vivid as normal vision', and rating 5 indicated 'no image at all, you only "know" that you are thinking of an object'. The VVIQ scores neither correlated with the *d'* values of the control condition, r(24) = .293, p = .165, nor with the hit rates of the blurred imagery condition, r(24) = .034, p = .875. But the VVIQ scores correlated with the hit rates of the scrambled imagery condition, r(24) = -.449, p < .05.

3.3 Discussion

The most important finding of Experiment 4 was that face imagery led to higher recognition rates for scrambled than for blurred faces. In the perception condition there was no difference between scrambled and blurred faces. This suggests that recognizing blurred and scrambled faces was equally difficult. However, the advantage found here for scrambled faces in the imagery condition suggests that top-down activation of faces predominantly primes featural representations. While in perception configural and featural processes are of comparable importance, mental imagery seems to activate featural more than configural representations. This assumption is further underlined by the correlation of the VVIQ scores and the hit rates of the scrambled imagery condition. The better the imagery abilities, as assessed with the VVIQ, the higher hit rates were for scrambled faces in the imagery condition.

Using introspection we may intuitively describe a mental image as blurred or fuzzy, as has accurately been described by Sartre in his important work The Imaginary (1940, trans. 2004). While trying to remember the face of his friend Pierre, Sartre finds that the face "is very imperfectly attained: some details are lacking (…) the whole is rather blurred" (p. 17). Only a photograph of Pierre can bring back to memory the featural details of the face. In contrast to this phenomenological description my findings suggest that introspection may misguide us in the search of the true nature that underlies visual mental images. Scrambled faces were recognized more accurately than blurred faces, indicating that rather than a blurred image we in fact activate relatively detailed featural representations when we imagine a face.

Perception and imagery apparently differ in the way that imagery only rarely involves any recognition, whereas perception always goes along with recognition. In this study I was interested whether a mental image of a face differentially primes configural and featural information and compared the results from the imagery condition with those from the perception condition. Insofar, in both the imagery condition and the perception condition participants had to match a name of a learned face with a presented face, albeit in the imagery condition they generated a mental image of the face before answering. Interestingly, there was no main effect of condition, neither for the hit rates nor for the RTs suggesting no overall facilitation through imagery. However, the significant interaction of condition and information indicates that mental images of faces do not activate configural representations as much as featural representations.

In contrast to Cabeza and colleagues (Cabeza et al., 1997) I found that imagery can indeed prime recognition of faces. But imagery essentially primes featural information: I found higher accuracy for scrambled faces in the imagery than in the perception condition. Blurred faces, however, were recognized equally well, whether or not a visual image of the face could be formed beforehand. Contrary to this study, Cabeza and colleagues did not differentiate between configural and featural representations. Had they included a task involving featural information, they might have found a priming effect of face imagery on face perception. Furthermore, their data analysis was restricted to response latencies. While the response times revealed no significant effects in the present study, response accuracy did. Finally, the tasks in the imagery and perception condition in Cabeza et al.'s study were inconsistent. In the present study the tasks were the same with the only difference being the mental image of the face, which was generated before the face stimuli were visually presented.

It has to be noted, however, that almost 70% of the blurred trials in the imagery condition were correctly recognized. I therefore do not claim that imagery fails to activate configural representations, but argue that when asked to spontaneously form a visual image of a newly learnt face people tend to activate featural more than configural information. Had participants been asked to specifically activate configural representations of a face (e.g., whose eyes are closer together, Peter's or David's) it is possible that configural representations could play a more important role.

The present results provide evidence that featural and configural processing can be differentiated in mental imagery. This goes along with findings revealed by means of neuro-imaging (Ishai et al., 2002). Moreover, the difference between featural and configural processing in face imagery may help to better understand inconsistent reports of prosopagnosic patients. Some people with prosopagnosia report no difficulties in forming mental images of faces they know (e.g. Bodamer, 1947; Pallis, 1955), while others report a disability to mentally visualize faces (e.g., Charcot & Bernard, 1883; Young et al. 1994). It may be possible that in the former group of patients the lesion affects only perception-driven activation of face representations while a top-down activation of featural representations is still possible. In the latter group, however, it is possible that top-down and perception-driven activation of face representations are equally affected by the lesion. It will have to be the issue of future brain-imaging studies with patients suffering from prosopagnosia with or without impaired imagery abilities to substantiate this proposition.

Another interesting issue is to discuss my findings against the background of studies on the verbal overshadowing effect. The term verbal overshadowing effect describes the phenomenon that people recognized faces less accurately when they previously described the face verbally (Dodson, Johnson, & Schooler, 1997; Schooler & Engstler-Schooler, 1990). Macrae and Lewis (2002) found that when participants adopt a local processing strategy (i.e., pay more attention to featural information), recognition of newly learned faces is impaired. Their finding suggests that not the verbal description per se hampers later recognition of faces, but the processing strategy adopted when describing a face. Describing a face verbally activates a local processing strategy, as faces are most often described by the features. The findings of Experiment 4 suggest that, similar to verbal description, a mental image of a face will activate featural representations. These findings could therefore have practical implications for criminal investigations when trying to find an offender based on the descriptions of eye witnesses. Because mental imagery of a face seems to mainly activate featural information, it will be the features that come to mind when witnesses are asked to remember the face of the person they saw committing a crime. Photofit pictures used by the police meet these concerns, as the faces are built up from different face parts. However, the verbal overshadowing effect suggests that an activation of the features later leads to impaired recognition of the face. Taking the findings of Macrae and Lewis together with my findings suggests that forensic psychologists have to be careful about the accuracy of the descriptions of witnesses and their ability to recognise the offender in a later line-up.

In conclusion, I found that although featural and configural processes can be separately activated in both mental imagery and perception, mental imagery seems to particularly activate featural representations. While performance in configural and featural trials was comparable in the perception condition, the importance of featural information was higher in face imagery. This finding suggests that imagery of faces activates featural representations more accurately than configural representations. In perception configural and featural representations played an equal role.

4. The Thatcher Illusion: Rotating the Viewer instead of the Picture

Abstract

Faces are difficult to recognize when presented upside down. This effect of face inversion was effectively demonstrated with the "Thatcher illusion" by Thompson (1980). It has been tacitly assumed that this effect is due to inversion relative to retinal coordinates. Here I tested whether this effect is due to egocentric (i.e. retinal) inversion or whether the orientation of the body respective to gravity also influences the face inversion effect. Using a 3D human turntable subjects were tested in five different body tilt (roll) orientations: 0°, 45°, 90°, 135°, and 180°. The stimuli consisted of 4 "normal" and 4 "thatcherized" faces and were presented in 8 different orientations in the picture plane. The subjects had to decide in a yes-no task whether the faces were "normal" or "thatcherized". Analysis of the d-prime values revealed a significant effect of stimulus orientation, and of body tilt. The significant effect of body tilt was due to a drop in d-prime values in the 135° orientation. This result is compared to findings of studies investigating the subjective visual vertical, where larger errors occurred in body tilt orientations between 90° and 180°. The present findings suggest that the face inversion effect mainly relies on retinal coordinates, but that in head-down body tilt orientations around 135° the gravitational reference frame has an arduous influence on the perception of faces.

4.1 Introduction

Visual objects are difficult to recognize when presented upside down. The effect of inversion differs depending on the type of visual stimuli, faces showing a significantly larger inversion effect than objects (Yin, 1969; for an overview see Valentine, 1988). A widely accepted explanation for this discrepancy goes back to the distinction between featural and configural information (e.g., Leder & Bruce, 2000). Featural information refers to information that is contained in the local parts (e.g., the individual shape of the nose); configural information refers to the spatial arrangement of the parts (e.g., the distance between the eyes and the mouth). In faces configural information plays a dominant role, while object recognition is much more based on local information contained in the features (Biederman, 1987; Marr, 1982; Tversky & Hemenway, 1984). Configural information has been shown to be more orientation sensitive than featural information (e.g., Leder et al., 2001; Nachson & Shechory, 2002; Searcy & Bartlett, 1996), and therefore face recognition is hampered when faces are presented upside-down. In this orientation, faces can only be recognized by matching their parts (Rock, 1973). Thompson (1980) effectively demonstrated that different orientation sensitivity of features and configurations may be responsible for the face inversion effect (FIE). He took a photograph of the former British prime minister and inverted eyes and mouth with respect to the whole face. Such a face looks extremely grotesque when viewed right-side up, but loses this grotesqueness when the face is inverted. This effect is now commonly referred to as Thatcher illusion. A possible explanation for this effect has been provided by Rock (1973). Recognition of such an inverted „thatcherized" face requires featural and configural information to be rotated mentally. Yet, the spatial transformation of all features and configurations overtaxes the capacity of the underlying mechanism (cf. Rock, 1973). Therefore it is difficult to mentally visualize what an inverted thatcherized face would look like right side up. The closer the orientation to upright the better the configural information can be extracted from the face. Numerous studies have been concerned with this phenomenon since then (e.g., Lewis, 2001; Rakover, 1999; Sturzel & Spillmann, 2000; Valentine & Bruce, 1985). Sturzel and Spillmann (2000) gradually turned different thatcherized faces from 0° through 180° and asked participants to report when the face switched from pleasant to grotesque, or vice versa. They found a relatively narrow changeover-zone between 97.2° and 118.3° where the change of expression occurred. Sturzel and Spillmann suggested that the striking change may be based on the step-tuning properties of hypothetical face neurons, rather than a gradual tuning curve. According to Sturzel and Spillmann face neurons respond best to faces in a tuning width of ±100° relative to the vertical. They claim that these face neurons may also respond to inverted faces, but inappropriately. In contrast, Lewis (2001) reported a gradual loss of configural information the further a face is turned away from upright. He recorded the reaction times of 40 participants while they discriminated thatcherized from normal faces which were presented in 10 different orientations. Based on these two studies, however, the nature of the dependence on the rotation-angle still remains equivocal.

Interestingly, almost all studies on the FIE have in common that they were conducted with upright observers. Therefore, it is often tacitly assumed that the effects induced by inverted stimuli are defined with respect to retinal coordinates. However, a stimulus can be upright or inverted with respect to retinal coordinates, or with respect to gravitational coordinates. Even though we are upright most of the time it is not evident that the reference frame underlying the FIE is of purely retinal origin. It is possible that it also depends on the orientation of the face stimulus with respect to the direction of gravity, since influence of extra-retinal information has been found in object recognition (Simons, Wang, & Roddenberry, 2002). In everyday life we mainly see faces in a gravitationally upright orientation, so it is plausible to assume that the direction of gravity can be implicitly encoded when faces are learned. In the upright body orientation, however, we are unable to disentangle the role of gravitational and retinal information because the two frames of reference are fully aligned. To investigate the influence of the gravitational frame of reference it is therefore inevitable to test participants not only in the upright body orientation but also when they are tilted.

The fact that body tilt can influence visual perception has already been demonstrated by Rock (1973). A square tilted 45° was no longer perceived as a square, but as a diamond. However, when the subject was tilted 45° and the square remained upright (resulting in roughly the same retinal image) subjects reported to see a square. Gaunet and Berthoz (2000) investigated the effect of gravity on the recognition of spatial environment. Their participants were tested upright and tilted 33° to the left and right. The task was to recognize photographed scenes, which were tilted in 15° steps from 0° to 90°. In contrast to Rock (1973) they found that gravity was only slightly important for recognizing scenes, and concluded that in their task it played no crucial role. Recent work by Clément and Eckhardt (2005) suggested that visual illusions such as the Ponzo illusion occurred less frequently when participants were lying on their side or supine compared to upright. In contrast, a study by Prinzmetal and Beck (2001) showed that the effect of visual illusions was in fact increased when observers were tilted 30°. Lipshits and McIntyre (1999) suggested a multi-sensory reference frame for the internal representation of visual stimuli. They showed a sequence of two lines of equal length which differed in orientation. The task was to memorize the orientation of the first line (reference line) and rotate the second line to the same orientation as the reference line, using a rotary knob. In the upright body orientation there was a clear advantage for reference lines that were horizontal and vertical (oblique effect). However, this preference disappeared when the lines were in fact presented retinally horizontal and vertical, but the participants were tilted 22.5° to the left or right. Using the same task under microgravity conditions (0g) the same preference as in upright body orientation was found (Lipshits, Bengoetxea, Cheron, & McIntyre, 2005). These findings suggest that visual stimuli may be stored in a multimodal frame of reference that includes information about gravity, but that in the absence of gravity the retinal reference frame suffices to determine the oblique effect. Buchanan-Smith and Heeley (1993) provided further evidence that the oblique effect cannot simply be explained by the retinal reference frame.

Moreover, tasks involving mental image transformations have also been shown to depend on body orientation. Corballis and colleagues (Corballis, Nagourney, Shetzer, & Stefanatos, 1978) tested participants in upright orientation, and when they were tilted 60° and 90° to the side and found that in a mental rotation task using alphanumeric and letter-like symbols as stimuli the gravitational reference frame indeed had an influence. Specifically, they found that "upright" is more aligned with the gravitational vertical than with the retinal vertical. Mast, Ganis, Christie, and Kosslyn (2003) investigated the performance in four different mental imagery tasks while participants were upright, horizontal, or supine. They found an influence of body orientation in two imagery tasks, suggesting that body tilt influences at least some processes associated with mental imagery.

Does body orientation have an influence on face recognition? On the basis of the findings reported above it can be hypothesized that the direction of gravity may also have an influence on the perception of faces. In particular, the visual illusions used in the studies by Prinzmetal and Beck (2001) and Clément and Eckardt (2005) are indeed visual stimuli involving predominantly configural processing. Since the FIE is based widely on the processing of configural information it is possible that it is also affected by body tilt. To my knowledge, only one study has yet been conducted to investigate the FIE in different body tilt orientations. Troje (2003) reported no changes depending on the body orientation and concluded that the retinal frame of reference is responsible for the FIE. However, this study only investigated observers that were upright and lying 90° on the side. Therefore, it seems premature to draw conclusions that are based on only one body tilt orientation. It is now important to study the FIE in a wider range of body tilt angles. In the present study I made use of the Thatcher illusion to investigate the FIE. I investigated whether body tilt influences the Thatcher illusion, or whether it can be fully explained by stimulus orientation with respect to retinal coordinates. Using different body tilts the gravitational and retinal frame of reference was disentangled. If indeed these the gravitational reference frame influences the Thatcher illusion, a differential effect of body tilt on stimulus orientation would be expected. If the Thatcher illusion is only based on retinal coordinates no effect of body tilt would be expected.

4.2 Experiment 5

4.2.1 Method

4.2.1.1 Participants

Thirteen participants ranging in age between 25 and 34 years voluntarily took part in this experiment. All but two participants reported to be right-handed. They could choose whether they wanted to be paid for participation or to receive course credits. All had normal or corrected to normal vision. Informed consent for participation was given prior to the experiment and the study was approved according to the Declaration of Helsinki (1991).

4.2.1.2 Stimuli

Four faces provided by the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany served as stimuli. The thatcherized stimuli were prepared using Adobe Photoshop©. With the elliptic tool, the eyes and mouth were cut out using a soft-contour feather of 5 pixels and were mirror-reversed round the horizontal axis. A sample stimulus can be seen in Figure 1. Then each face was rotated in the picture plane into the 8 different angles of stimulus orientation (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°).

In a pilot study a separate group of nine participants was tested outside the turntable in upright body orientation using the upright (0°) and inverted (180°) stimuli described above. The effect of stimulus orientation for accuracy was significant, $F(1,4) = 9.765$, $MSE = 4.925$, $p < .05$, showing that inverted thatcherized faces were not detected as accurately as upright thatcherized faces. The result of this pre-test demonstrated that the stimuli are appropriate for testing the FIE.

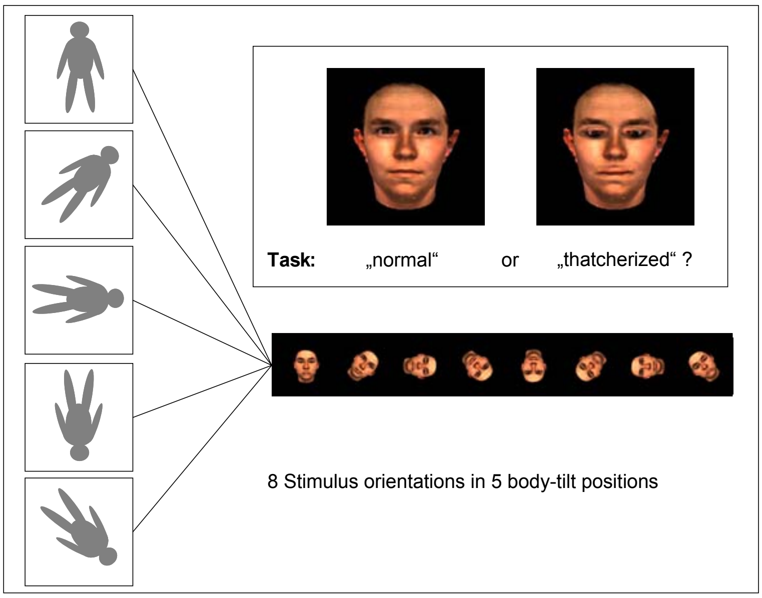

Figure 10: Design and sample stimulus: In each of 5 body tilt orientations 4 thatcherized and normal faces were presented in 8 stimulus orientations.

4.2.1.3 Apparatus

The participants were tested in five different body tilt orientations (roll), 0° (upright), 45°, 90° (horizontal, right ear down), 135°, 180° (upside down).

The face stimuli appeared in eight different orientations (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°). Thus, 8 x 5 combinations were possible per stimulus. Four different individual faces were used which appeared in a thatcherized and a normal version. In total, 320 different trials were applied (8 x 5 x 8). The experiment was run using a 3D human turntable (Acutronik, Jona, Switzerland) at the Department of Neurology, University Hospital Zurich. The turntable consisted of three servo-controlled motor driven axes which could be separately controlled.

The participants were seated on a chair mounted on the 3D human turntable and firmly secured with safety belts. Participant's naso-occipital axis was aligned with the centre of rotation. The head was restrained with a thermoplastic mask (Sinmed BV, Reeuwijk, The Netherlands), which was individually moulded for each participant. The mask was attached to the back of the chair ensuring effective restraint of the head without discomfort. This fixation in combination with the belts ensured a stable position in head-down body tilts. The participants were brought to one of the four body tilt orientations (45°, 90°, 135° or 180°) with a speed of 45°/s and an acceleration of $45°/s^2$. After a delay of 2 s the participants were prompted to start the first trial by pressing one of the response buttons. Jaggi-Schwarz and Hess (2003; personal communication with Hess) found no torsional nystagmus (VOR) two seconds after stopping the body rotation at this speed. Therefore, this interval is long enough for vestibular driven eye movements to dissipate, which could interfere with the perceptual encoding of the face stimuli. The stimuli were presented via a Macintosh G3 Powerbook which was mounted onto a frame attached to the chair, using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). The participants responded by using a PsyScope button box which was attached to the frame. Reaction times and button presses were recorded.

4.2.1.4 Design

A within-subjects design was used. Participants completed two blocks of 32 trials in each body orientation, encompassing all stimulus orientations and faces, but each face appeared either normal or thatcherized. Whether a face was thatcherized or normal was pseudo-randomized with the constraint that half of the stimulus orientations of each face were thatcherized and half were normal. Thatcherization was counterbalanced within-subjects between blocks. The order of body tilt orientations was attained as follows: Four random orders of the five body orientations were created. Using Latin squares, five orders were generated from each random order. Thus, 20 orders were computed. Each participant underwent two orders of body orientations. The order of the trials in each block was randomized online.

4.2.1.5 Task and Procedure

The participants were given written and oral instructions. The task was to decide whether a face was "normal" or "thatcherized" by pushing the corresponding key on the response box. The participants were tilted into one of five body orientations and were then presented with the first test face. Each test face was presented for 200 ms in one of eight stimulus orientations; either normal or thatcherized and the participants had to respond as fast and as accurately as possible. After each block the participants were brought back to the upright body orientation and were able to take a rest. The length of the break was self paced, but the minimum duration was 30 s. As soon as the participants were ready they were tilted into the next body orientation and the experiment continued with the next block.

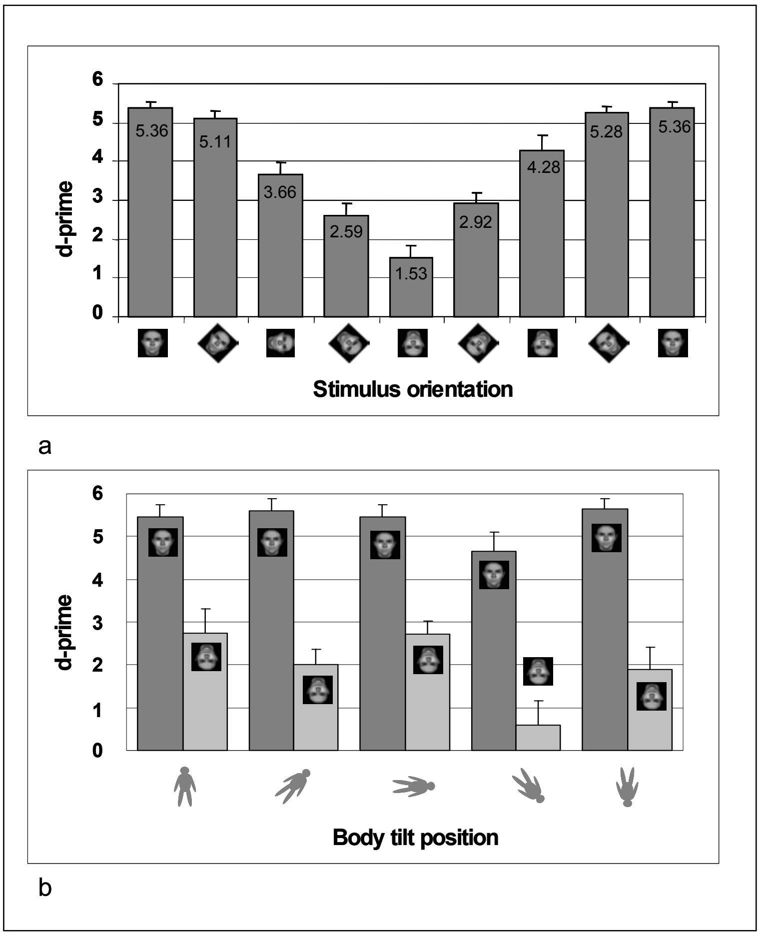

Figure 11: a) Main effect of stimulus orientation. Mean d-prime values of all stimulus orientations independent of body tilt. Error bars depict standard errors of the mean (SEM). b) Main effect of body tilt. Mean d-prime values of all body tilt orientations for retinally upright and inverted stimuli. Error bars depict standard errors of the mean (SEM).

4.2.2 Analysis

D-prime values (*d'*) and reaction times (RTs) of the correct responses were analyzed. Less than 0.8 % of the trials were treated as outliers and were excluded from analysis because RTs were above 3000 ms. D-prime values were calculated for each subject by subtracting the z-transformed false alarm rate from the z-transformed hit rate. First, 5x8 analyses of variance (ANOVA) were run including all body tilt orientations and all stimulus orientations. In addition, to investigate the FIE 5x2 analyses of variance (ANOVA) were conducted with all five body tilts and upright and inverted face stimuli as within subjects' factors.

4.2.3 Results

The mean *d'* values for each body tilt orientation and stimulus orientation are illustrated in Figure 2. Figure 2a illustrates the *d'* values of all stimulus orientations, independent of body tilt, Figure 2b shows the effect of body tilt on *d'* for upright and inverted stimuli. The 5x8 ANOVA on the *d'* values revealed a significant effect of body tilt, $F(4,48) = 6.307$, $MSE = 1.346$, $p < .001$ and of stimulus orientation, $F(7,84) = 39.481$, $MSE = 3.285$, $p < .001$, and a significant body tilt x stimulus orientation interaction, $F(28,336) = 1.751$, $MSE = 2.238$, $p < .05$. To specifically investigate the FIE, I computed a 5x2 ANOVA on the *d'* values of retinally upright and retinally inverted stimuli in all 5 body tilts, which revealed significant main effects of body tilt, $F(4,48) = 6.24$, $MSE = 1.56$, $p < .001$, and stimulus orientation, $F(1,12) = 85.10$, $MSE = 4.33$, $p < .001$. The interaction (body tilt x stimulus orientation) did not reach statistical significance ($p = .222$). Post-hoc pairwise comparisons (Bonferroni corrected) revealed that the main effect of body tilt was due to the tilt-angle of 135°. Here, participants had lower *d'* values. Only comparisons involving the 135° orientation reached statistical significance: the comparison between 0° and 135° revealed a significant difference (p < .05), and so did the comparison between 90° and 135° (p < .01). The comparison between 45° and 135° reached marginal significance (p = .11), and the comparison between 180° and 135° did not reach statistical significance (p = .304).

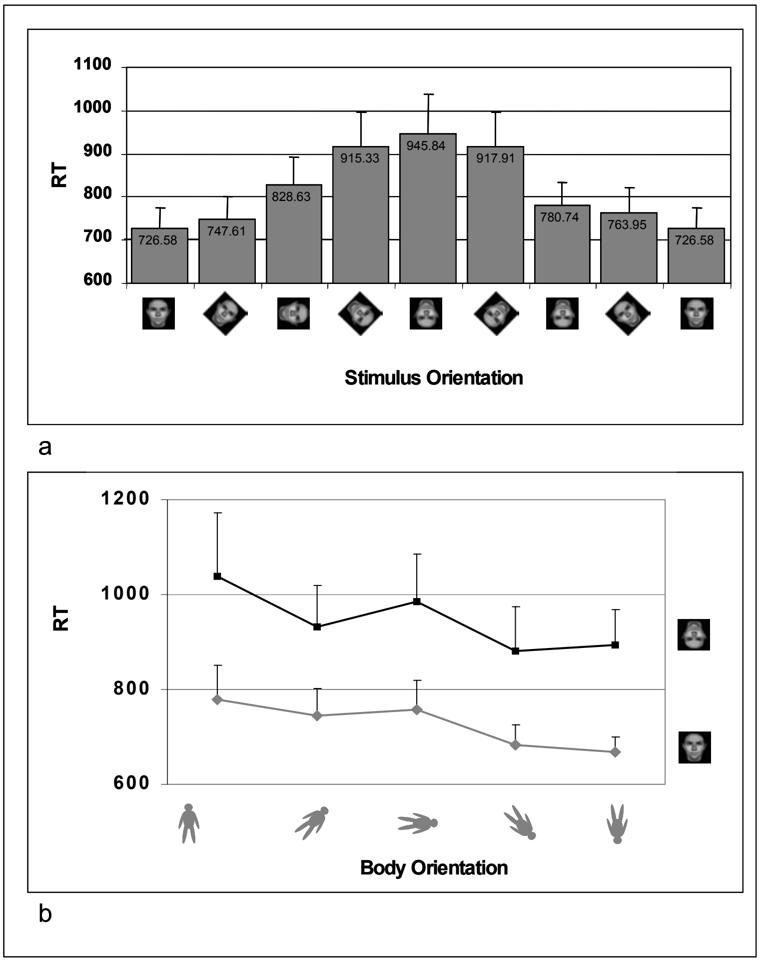

Figure 12: a) Main effect of stimulus orientation. Mean RTs of all stimulus orientations independent of body tilt. Error bars depict standard errors of the mean (SEM). b) Main effect of body tilt. Mean RTs of all body tilt orientations for retinally upright and inverted stimuli. Error bars depict standard errors of the mean (SEM).

The mean RTs are shown in Figure 3. Figure 3a illustrates the RTs of all stimulus orientations, independent of body tilt. Figure 3b shows the effect of body tilt on the RTs for retinally upright and inverted stimuli. The 5x8 ANOVA on the RTs revealed a main effect of stimulus orientation, $F(7,84) = 16.709$, $MSE = 29277$, $p < .001$, but neither the main effect of body tilt nor the interaction body tilt x stimulus orientation reached statistical significance.

The 5x2 ANOVA revealed a main effect of body tilt, $F(4,48) = 3.6$, $MSE = 22528$, $p < .05$. The more participants were turned away from upright the faster they responded. It is worth noting that d-prime values did not decrease in the upside-down body orientation, and therefore the shorter RTs were not due to a speed-accuracy trade off. The effect of stimulus orientation also reached statistical significance, $F(1,12) = 18.19$, $MSE = 85879$, $p < .01$, but the interaction (body tilt x stimulus orientation) was not significant ($p = .69$).

4.3 Discussion

Three findings of this experiment deserve special attention. First, the inversion effect was generally based on retinal coordinates. Second, when the body was tilted 135°, thatcherized faces were more difficult to detect. Third, the further observers were tilted away from upright the faster they detected retinally inverted and retinally upright thatcherized faces.

Faces were more difficult to process when they were retinally inverted than when they were retinally upright. In fact, it took participants more time to discriminate thatcherized from normal faces the more they were rotated away from retinal upright. This finding is in line with the results of Lewis (2001) who reported a gradual increase of RTs with increasing stimulus orientation of thatcherized faces. Whereas Lewis (2001) only reported an increase of RTs I also found a decrease of d-prime. This finding indicates that configural information gradually is hampered the further a face is turned away from retinal upright. The data of Experiment 5 therefore contradict findings of Sturzel and Spillmann (2000), who reported a relatively narrow range of stimulus rotation angle where a thatcherized face loses its grotesqueness. They suggested that step-tuning properties of face neurons may be responsible for their results. The increasing RTs for rotated faces make it more likely to suggest a mental rotation process that underlies the findings from this study. Observers had to mentally rotate each face to a retinal upright orientation, and the time to perform this process increased with angle of stimulus orientation; at the same time the accuracy decreased.

Yet another important finding is that the angle of body tilt had an influence on the detection of the Thatcher illusion. In particular, at 135° it was more difficult to detect the changes in the faces. This finding suggests that gravitational direction indeed has an influence on the FIE. It is noteworthy that *d'* in the upside down body orientation (180°) did not differ from any of the other body orientations, and thus the results are not explainable by a general decline in performance caused by head-down orientations. De Schonen et al. (1998) studied the FIE in microgravity and found no change when compared to performance on the ground. In microgravity, however, there is no sensory information regarding the direction of gravity and participants rely exclusively on visual information. Similarly, Troje (2003) reported no effect of body orientation on the FIE. However, his study was confined to the use of two different body tilt orientations only. When only looking at 0° and 90°, my data confirm the findings of Troje (2003). Including a wider range of body tilt orientations revealed that the direction of gravity can influence the FIE. Gaunet and Berthoz (2000) tested the influence of gravity in a natural scene recognition task. They also only used one small body tilt (i.e., 33°). They concluded that gravity is not a crucial factor in their experiment. The present findings suggest that an effect of gravity may indeed have been found in body tilts around 135°. For natural scenes the effect of gravity may even exceed the effect I found for faces, because – unlike faces – natural scenes always appear gravitationally upright.

What could be the reason for the distinctive decline in performance at 135°? In this context it is interesting to note that studies investigating the subjective visual vertical report largest errors in body tilt orientations around 135° (e.g., Kaptein & Van Gisbergen, 2004; Schöne, 1964; Udo de Haes, 1970; Van Beuzekom & Van Gisbergen, 2000). Interestingly, not only the deviation from the physical vertical but also the variance of the subjective visual vertical reached its maximum in head-down body tilts between 120° and 150° (Mast, 2000; Mittelstaedt, 1999). This indicates that the participants have less reliable reference information for the perception of the vertical. As a consequence, the participants have difficulties in judging the orientation of visual stimuli with respect to gravity. The retinal and gravitational references are not aligned in 45°, 90° and 135°, but in 135° the deviation between the retinal up and the perceived gravitational up is largest. Here, the two references deviate by more than 90° (this is also true for the upside down orientation but there the reference frames are again aligned, albeit in exactly opposite directions). This disparity may result in a reduced confidence in the spatial reference information underlying the FIE and thus finally disrupts task performance.

I assume that the FIE is explainable with retinal coordinates as long as the retinal and gravitational frame of reference do not deviate substantially (i.e., deviations larger than 90°). No such conflict arises in the upside down orientation where the reference frames are again perfectly aligned but point in opposite directions. Therefore, the FIE is essentially the result of visual information processing and only in head-down orientations around 135° the extra-retinal reference information will unfold its effect. In the present study however, only roll orientations were tested. It will have to be the aim of future studies to investigate whether rotations round other body axes will produce a similar effect (e.g., head down tilts in the body pitch direction). The overall interaction of body tilt and stimulus orientation further suggests that body tilt has an influence on the perception of orientation-sensitive stimuli such as thatcherized faces. Thus, this study demonstrated that a gravity-based component exists and can interfere with task performance.

Finally, it seems that “standing on one’s head” speeds up detection of retinally upright and inverted thatcherized faces. Taking the d-prime values into account it is unlikely that participants just wanted to get over more quickly with the head-down body conditions, as the detection ability of thatcherized faces did not decrease in body tilt orientations of 180°. Furthermore, this advantage of being upside-down only applied for retinally upright and inverted faces. Whether this finding was specific to the task used in this study and why it only occurred for upright and inverted faces will have to be the issue of further studies. As the participants in this study stayed in one body tilt orientation for approximately 60 s maximally, nothing can be said about the time course for longer time periods. This enhancement may eventually decline after a certain time, presumably when the cardio-vascular system has regularized the blood pressure.

In conclusion, Experiment 5 has shown that the Thatcher illusion is based mainly on the orientation of the face stimulus with respect to the retinal reference frame. However, in head-down body tilt orientations around 135° the gravitational and retinal reference frames deviate substantially and as a consequence participants have difficulties in unambiguously perceiving the orientation of visual stimuli. This effect was found for faces, whether it also applies for other complex visual stimuli will have to be the issue of future studies.

5. Is the Thatcher Illusion Restricted to Faces?

Abstract

Faces are very sensitive to inversion, which has been effectively demonstrated in the Thatcher illusion, where eyes and mouth were inverted within the whole face (Thompson, 1980). When viewed in an upright orientation such a face looks extremely grotesque, but loses this grotesqueness when turned upside down. Here I investigate whether the same effect can be found for houses. Analogous to thatcherized faces, I inverted windows and doors within whole houses and presented thatcherized houses and faces for 250 ms each to 16 participants. Their task was to indicate whether the house or face was normal or thatcherized. An ANOVA on the *d'* values revealed a significant effect of stimulus type, orientation and a significant interaction of stimulus type and orientation. Post-hoc two-sample t-tests however revealed that inverted thatcherized houses were not detected less accurately than upright thatcherized houses. For faces I found the expected Thatcher effect. These findings suggest that the Thatcher effect might be unique for faces.

5.1 Introduction

Inverted faces are more difficult to recognize than inverted pictures of other object classes (Valentine, 1988; Yin, 1969). That faces are indeed sensitive to inversion was effectively demonstrated in the Thatcher illusion (Thompson, 1980). Thompson (1980) inverted the eyes and mouth within the face of former British Prime Minister Margaret Thatcher. When viewed in an upright orientation her facial expression appears extremely grotesque, however this grotesqueness disappears when the face is turned upside down. A widely accepted explanation for this asymmetry is that configural information is believed to be particularly important in face processing (Tanaka & Farah, 1993; Farah et al., 1995; Tanaka & Sengco, 1997), while objects are processed part-based (Biederman, 1987; Marr, 1982; Tversky & Hemenway, 1984). Configural information has been considered to be much more orientation sensitive than featural information (Leder et al., 2001; Maurer et al., 2002; Rock, 1973; Searcy & Bartlett, 1996; Farah et al., 1995; Yin, 1969), hence a main source of information is disrupted when turning a face upside down.

For this reason, the Thatcher-effect has been traded as face specific. However, Rock (Rock, 1988) mentioned a similar effect when inverting individual letters within a whole word. Still, this effect is not as striking and allowedly not as strong as found in thatcherized faces. Rock's explanation for this illusion is that an inverted image has to be mentally rotated to the upright position in order to perceive the expression on the face. Inverted faces overtax mental rotation mechanisms and can therefore only be processed part by part (Rock, 1973, 1988). And because parts are not sensitive to inversion, one does not succeed in noticing that these parts are actually inverted with respect to the whole image. A slightly different explanation for the Thatcher illusion is provided by Rakover (1999). He argues that the whole face is more dominant than the individual features and interprets the strangeness of upright thatcherized faces as a result of our inability to grasp the eyes as locally inverted in the pattern of the upright whole face. When inverted, the thatcherized face is perceived as an inverted face and the eyes are perceived within this frame in their true orientation: upright eyes. But in this inverted case, the whole face is perceptually not as dominant as the individual features, because inversion disrupts perception of the whole face (see also Farah et al., 1995).

Various studies have been conducted scrutinizing the nature of the Thatcher illusion. For example, the dependence of the rotation angle has been investigated. Lewis (2001) recorded the reaction times of participants while they discriminated thatcherized from normal faces which were presented in 10 different orientations and found a general increase the further away faces were turned from upright. In Experiment 5, I found very similar results not only for reaction times, but also for *d'* values, indicating that configural processing is gradually decreased, the further a face is turned away from upright (see also Lobmaier & Mast, in press). In contrast, Sturzel and Spillmann (2000) suggested that step-tuning functions of hypothetical face neurons may be responsible for this effect. They gradually turned different thatcherized faces from 0° to 180° and participants were required to indicate when the face switched from pleasant to grotesque. The results suggested a relatively narrow change-over zone between 97.2° and 118.3° where the change of expression occurred rather than a gradual tuning curve (Sturzel & Spillmann, 2000). According to these authors, face neurons may respond inappropriately to inverted faces, resulting in the Thatcher illusion.

In the present study I investigate whether the Thatcher effect also applies to other stimuli that are made up of different parts, but where the configuration of these parts is crucially important, similar to the interrelationship of parts and configuration in faces. Houses seem to meet these premises, as they are all made up of a certain number of parts (e.g., windows, doors, etc.) which are individually arranged. I compared the perception of upright and inverted thatcherized faces with upright and inverted thatcherized houses, to ascertain whether the Thatcher effect is unique for faces.

5.2 Experiment 6

5.2.1 Method

5.2.1.1 Participants

16 Participants (8 female, 8 male) ranging in age from 22 to 49 (mean = 30 years) voluntarily took part in the experiment and were naïve to the goal of the Experiment. Four participants reported to be left handed. All reported normal or corrected to normal vision.

5.2.1.2 Apparatus

The Experiment was run on a 15.1” Pentium 4 portable Computer using Superlab Pro 2.0.2 running on Windows NT. The experiment took place in a dimly lit room. The participants were seated on a height-adjustable chair and responded by using a Cedrus© Response Pad (RB-520). They were required to keep the head still on a head-rest, keeping a viewing distance of 500 mm constant. Participants underwent a total of 10 blocks, 5 blocks with houses and 5 blocks with faces. Each face stimulus appeared 45 mm wide and 56 mm high and thus subtended a visual angle of approximately 5.2° horizontally. The faces were placed on a black square (67 x 67 mm) in the centre of the screen. The house stimuli were approximately 62 mm wide and 35 mm high and were placed on a grey rectangular (47 x 67 mm) in the centre of the screen. The house stimuli subtended to a visual angle of approximately 7.1° horizontally.

5.3.1.3 Stimuli

Five faces and houses served as stimuli for this experiment. The faces were provided by the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany. Thatcherized faces were created by inverting the eyes and mouth within these faces (Thompson, 1980). The house stimuli were taken from the house-database of Alumit Ishai, Institute of Neuroradiology, University of Zurich, Switzerland. To accomplish “thatcherized” houses, the entrance, windows and garage door were inverted within the house. All stimuli appeared either upright or inverted.

5.3.1.4 Task and Procedure

Each trial began with the appearance of a fixation cross in the centre of the screen, followed by a stimulus. Each stimulus was presented for 250 ms and task was to decide as accurately and fast as possible whether the stimulus was normal or thatcherized. Participants answered with the preferred hand for normal stimuli and with the subdominant hand for thatcherized stimuli. After pressing one of the answer-buttons the fixation cross reappeared and was presented until participants initiated the next trial by pressing the centre button. The inter-trial interval was therefore self-paced; the next stimulus appeared 500 ms after initiating a new trial with the centre button.

Prior to the experiment participants gave informed consent. Then they received written and oral instructions which were followed by 16 practice trials, encompassing all experimental conditions to make sure they understood the task. None of the stimuli that appeared in the practice trials were later used in the experiment proper. In the experiment proper participants underwent 5 blocks with 20 face stimuli (5 different faces x 2 thatcherized or normal x 2 upright or inverted) each and accordingly, 5 blocks with 20 house stimuli each. House blocks alternated with face blocks; which block the experiment began with was counterbalanced across participants.

5.2.2 Results

D-prime values and reaction times (RTs) of correct responses were analyzed. D-prime values were calculated for each participant and condition by subtracting the z-transformed false alarm rates from the z-transformed hit rates. A hit was defined as a correctly detected thatcherized stimulus; a false alarm was defined as a stimulus that was normal, but mistakenly perceived as thatcherized. The mean d-prime values are shown in Figure 1. The mean *d'* value for upright faces was 4.01 (*SD* = 1.198), for inverted faces 1.16 (*SD* = 1.214), for upright houses the mean d-prime value was 1.65 (*SD* = 1.01) and 1.19 (*SD* = 0.89) for inverted houses. As the independent samples t-test revealed no effect of group, the data of the two groups was pooled for the subsequent analyses. A 2 x 2 analysis of variance (ANOVA) was carried out, with stimulus (face, house) and orientation (upright, inverted) as within subjects factors, revealing a significant main effect of stimulus, $F(1,15) = 29.836$, $MSE = 0.73$, $p < .001$, and of orientation, $F(1,15) = 222.411$, $MSE = 0.197$, $p < .001$. The interaction stimulus x orientation was also significant, $F(1,15) = 26.878$, $MSE = 0.847$, $p < .001$. A post-hoc two-sample t-test (two-sided) comparing upright with inverted stimuli revealed that upright thatcherized

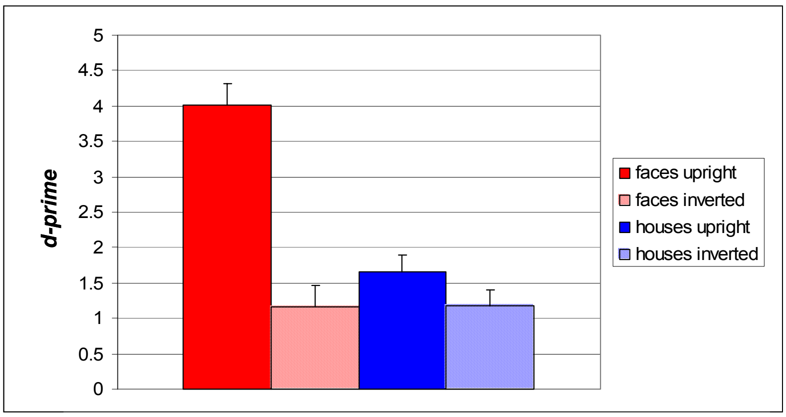

Figure 13: Mean d' values for upright and inverted thatcherized faces (red) and upright and inverted thatcherized houses (blue). Error bars depict standard errors of the mean (SEM).

faces were detected significantly better than inverted thatcherized faces, $T = 10.391$, $p < .001$. The t-test with upright and inverted houses however did not reach statistical significance, $T = 1.971$, $p = .068$.

The mean RTs for thatcherized stimuli were as follows: for upright thatcherized faces 685 ms (SD = 179.18 ms), for inverted thatcherized faces 1013.88 ms (SD = 326.63 ms), for upright thatcherized houses 1084.94 ms (SD = 287.81 ms) and for inverted thatcherized houses 1176.95 ms (SD = 414.4 ms). For normal stimuli the RTs were 766.27 ms (SD = 328.67 ms) for upright faces, 1001.13 ms (SD = 417.35 ms) for inverted faces, 965.79 ms (SD = 388.02 ms) for upright houses, and 1165.17 ms (SD = 406.9 ms) for inverted houses. A 2 x 2 x 2 ANOVA with the factors thatcherization (that), stimulus type (stim) and orientation (ori) was carried out on the reaction times, revealing a significant effect of stimulus, $F(1,15) = 22.735$, $MSE = 75527.5$, $p < .001$, and orientation, $F(1,15) = 28.599$, $MSE = 51136.7$, $p < .001$, but no significant effect of thatcherization. Whereas the interaction stimulus * orientation was significant ($F(1.15) = 7.765$, $MSE = 19104.9$, $p< .05$) all other interactions failed to reach statistical significance. Post-hoc t-tests comparing the RTs of inverted and upright thatcherized houses and faces revealed that upright thatcherized faces were detected faster than inverted thatcherized faces. However, there was no significant difference between upright and inverted houses.

5.3 Discussion

Thatcherized house stimuli were created by inverting entrance, windows and garage door within the whole house, similar to thatcherized faces and the ability to detect upright and inverted thatcherized houses and faces was compared. For faces the Thatcher-effect was generally larger than for houses. This finding is reflected in the significantly higher d-prime values and lower RTs for upright faces as compared to inverted faces. Upright houses were detected neither more accurately nor faster than inverted houses.

This finding suggests that the Thatcher illusion might be unique for faces. This is interesting, because both houses and faces are made up of different components which are distinct in their configural organisation. The Thatcher illusion is a result of the mutual interrelation of featural and configural information on the one hand and stimulus orientation on the other. Although most houses share common basic parts (e.g., doors, windows, etc.) which are configurally organized, the present data suggest that the interrelationship of featural and configural information contained in houses behaves differently than the interrelationship of features and configuration of faces. Insofar, it seems that face processing has aspects that are unique and do not apply for processing of other objects. However, Diamond and Carey (1986) have suggested that what makes faces unique is the expertise human beings have with them. They found a similar inversion effect when dog breeders looked at photographs of dogs of their breed of expertise. Specifically they claimed that a similar processing strategy is employed for objects we know well as for faces (see also e.g., Gauthier et al., 2000; Gauthier & Tarr, 1997). It is therefore questionable whether for example architects would have shown an effect with thatcherized houses. However, on the basis of the present data it can be said that houses are processed mainly part-based and that configural information plays a minor role. This is reflected in the fact that there was no effect of inversion for houses, neither for the *d'*prime values, nor for the RTs. Further support for this assumption lies in the finding that even upright thatcherized houses were not detected as accurately as upright thatcherized faces. This suggests that in upright as well as inverted houses featural information plays a dominant role, thus we do not notice the strangeness of inverted windows in an otherwise upright house.

Whereas there are numerous studies that show how faces are more prone to inversion than other objects (e.g., Diamond & Carey, 1986; Farah et al., 1995), to my knowledge nobody has investigated the effect of other objects that have been thatcherized by inverting certain parts relative to the whole array. The common finding that inverted objects are recognized more accurately than inverted faces (e.g. Yin, 1969), is reflected in the present data in the interaction of stimulus type and inversion. With the present experiment I found direct evidence that other objects such as houses are processed predominantly on the basis of parts.

6. Are featural and configural face processing strategies dissociable? Evidence from an fMRI Study

Abstract

In this study event related fMRI was used to explore processing mechanisms of featural and configural face information. In a delayed matching-to-sample task subjects decided whether an intact test face matched a precedent scrambled or blurred cue face. By means of dividing a face into its constituent parts and scrambling these parts the face loses its configural information. Featural information is destroyed by means of blurring. The test face induced differential neural activation depending on what information the cue face contained. Test faces following scrambled cue faces evoked increased activation in the left fusiform gyrus, left parietal lobe and left lingual gyrus, compared to test faces following a blurred cue face. Test faces following after blurred cue faces evoked increased activation bilaterally in the middle temporal gyrus, compared to test faces following scrambled cue faces. Consistent with previous behavioural studies the results of this experiment suggest that processing featural and configural face information are mediated by dissociated neural systems. This challenges the view that faces are processed as unparsed wholes. Rather the findings of this study suggest that featural and configural information is extracted and processed following distinct neural pathways before it is combined into a whole representation of the face.

6.1 Introduction

The processes underlying human face recognition have been the subject of numerous behavioural and neuro-imaging studies. Most neuro-imaging studies found a brain region in the fusiform gyrus which seems to respond specifically to human faces, termed fusiform face area (FFA) (Grill-Spector, Knouf, & Kanwisher, 2004; Haxby et al., 2001; Ishai, Ungerleider, Martin et al., 2000; Ishai et al., 1999; Kanwisher, 2000; Kanwisher et al., 1997; Kanwisher et al., 1999; Rossion et al., 2003). Many authors claim that, as opposed to objects, faces seem to be processed configurally, instead of part based (Diamond & Carey, 1986; Farah et al., 1995; Farah et al., 1998; Tanaka & Farah, 1993). The configuration of a face is understood as the information contained in the spatial interrelationship of its features. Nevertheless, recent studies have pointed out that featural information also plays a role in the processing of faces (Cabeza & Kato, 2000; Schwaninger et al., 2002). Further findings suggest that configural and featural information are processed following separate pathways (Bartlett et al., 2003; Cabeza & Kato, 2000; Rossion et al., 2000; Schwaninger et al., 2002). In a PET study, Rossion and colleagues (2000) found hemispheric differences when their subjects attended to featural or configural information. When faces had to be matched according to their configuration, the right middle fusiform gyrus showed more activation than the left homologous region. In part-based processing the activation in the right middle fusiform gyrus was reduced, but enhanced in the left middle fusiform gyrus. For objects, no such double dissociation could be found in these face specific regions. However, for objects hemispheric asymmetric activation for configural and featural information was found in the inferior parietal lobe and in the superior temporal gyrus (Weissman & Woldorff, 2005) and in the occipito-temporal regions (Martinez et al., 1997). In a study with patients with unilateral right or left lesions centred in temporal-parietal regions Robertson et al. (1988) found an asymmetry for local and global features (Robertson, Lamb, & Knight, 1988). Patients with right hemisphere lesions showed better performance in processing local features, patients with lesions in the left hemisphere performed better when processing global features. In the present study I scrutinize whether configural and featural face processing mechanisms can be dissociated. In the study by Rossion and colleagues (2000) subjects had to attend to either featural or configural information in a block design. The task was to match face pairs. These face pairs either differed in the spacing of the features (configural block) or in the features themselves (featural block). The subjects knew in each block what information they had to look for. Therefore their results could be the effect of different attention strategies, rather than different processes related to the stimuli. In the present study a delayed matching to sample task was used applying an event related design. Subjects first saw either a scrambled or a blurred face (cue face) and they had to decide whether a subsequent intact face (test face) was the same or not. In scrambled faces global configural information is reduced while local featural information remains intact. In blurred faces the detail information of the features is hampered while the overall configuration of the face is unrestricted (Schwaninger et al., 2002). By keeping the test face intact, the visual input of the critical stimulus remained the same. What changed was the cued information. Participants could only solve the task by using either configural or featural information, depending on the cue face. I expected to find differential activation depending on whether a scrambled or a blurred face preceded the test face. According to Rossion and colleagues (Rossion et al., 2000) a differential activation would be expected within the FFA. Alternatively, if the FFA in fact does reflect the face recognition unit (Bruce, 1988; Schwaninger et al., 2002) but is responsible for more holistic aspects of face processing, a differentiation between configural and featural processes differential activation could be expected earlier in the visual pathway. Similar to the model described elsewhere (Schwaninger et al., 2002) it could be conceivable that featural and configural information is first processed following two distinct pathways and is then combined into a more holistic face representation in the FFA. If so, different assumptions can be made about these pathways. On the one hand it could be expected that configural (metrical) information is processed via dorsal pathways, whereas featural information is processed via ventral pathways analogous to the "what" and "where" system (Haxby et al., 1991; Ungerleider & Mishkin, 1982). On the other hand a hemispherical difference could be expected for featural and configural processing (Martinez et al., 1997; Rossion et al., 2000). Yet another possibility is a combination of the two propositions, namely that featural information is processed more left-lateralized and ventrally, and configural information is processed in the right hemisphere and dorsally. Finally, if faces were processed purely holistically (i.e., featural and configural information is not processed following distinct pathways) no differential activation would be expected between test faces following scrambled faces and test faces following blurred faces. Differential activation in blurred and scrambled trials, however, would suggest differential processing of featural and configural information.

6.2 Experiment 7

6.2.1 Method

6.2.1.1 Participants

Fourteen right-handed subjects ranging in age between 24 and 32 years (mean 27.1 years) took part in this study. All gave informed consent and were treated according to the declaration of Helsinki. All subjects were paid for their participation at the end of the experiment.

6.2.1.2 Stimuli

Blurred, scrambled and intact faces were used as stimuli. In the control condition lines in four different orientations were used, either placed on a black background or on an array the same size as the faces. This array was a special scrambled version of a stimulus face, where an intact face was cut into small parts and rearranged so that it contained no featural and no configural information. Examples of the stimuli can be seen in Figure 1. The stimulation was presented via MR-compatible video goggles (MAVision 2000 fMRI, Resonance Technology, Inc.).

6.2.1.3 Task

The experiment was conducted using Presentation (www.neurobs.com). A trial started with a fixation cross, which was presented during 3 seconds. In the experimental conditions either a blurred (cueblr) or a scrambled face (cuescr) was presented for 5 seconds, followed by fixation cross (5000 ms) and an intact test face. The test face disappeared after 2000 ms or as soon as the subjects responded. Depending on whether a scrambled or blurred face preceded this test face was coded testscr, or testblr, respectively. The task was to decide whether the intact face was the same person shown

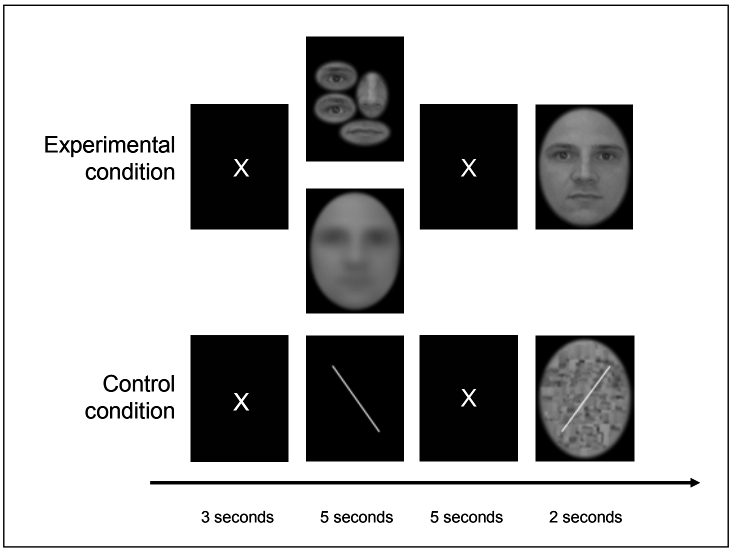

Figure 14: Design. After a fixation cross either a scrambled or blurred face (experimental condition) or a line (control condition) was presented for 5 seconds (cue). After a delay an intact face or control stimulus was presented (test stimulus). The task was to decide whether the test stimulus was the same as the cue stimulus.

in the cue face. In the control condition a line appeared instead of the cue face. Instead of the test face another line was presented on the scrambled array described above. Subjects had to decide whether the two lines had the same orientation. Thus, the test stimuli in the control condition contained virtually the same perceptual information as the test faces and the control task was a discrimination task as was the task in the experimental conditions. Subjects responded by pressing a button with the right index finger for same stimuli and with the left index finger for different stimuli. The procedural order of a trial can be seen in Figure 14.

6.2.2 Magnetic Resonance Imaging and fMRI Data Analysis

Gradient echo, echoplanar imaging was performed using a GE Signa 3 Tesla scanner, obtaining volumes of 32 3.5 mm thick axial images which were recorded in an interleaved manner (TR = 2.4 seconds, TE = 32 ms, FA = 90, FOV = 26 cm, 96 x 96 matrix). Two runs consisting of 306 volume scans each were obtained employing an event related design. The scans were aligned along the AC/PC axis and were then processed and analysed using SPM2 (Wellcome Department of Cognitive Neurology, London). To correct for their different acquisition times, the signal measured in each slice was shifted relative to the acquisition time of the first slice using a sinc interpolation in time. The images of each subject were realigned to the first image to correct for head movement. Then the images were normalized into stereotaxic anatomical Montreal Neurological Institute (mni) space by using the transformation matrix calculated from the first volume of each subject and the EPI template provided by SPM2. Afterwards, the normalized data with a resliced voxel size of 3x3x3 mm were smoothed with a Gaussian kernel (full-width at half-maximum 6 mm) to accommodate intersubject variation in brain anatomy. All analyses were restricted to trials on which responses were correct. The expected hemodynamic response at stimulus onset for each event-type was modelled by two response functions, a canonical hemodynamic response function (HRF) (Friston et al., 1998) and its temporal derivative. The temporal derivative was included in the model to account for the residual variance resulting from small temporal differences in the onset of the hemodynamic response, which is not explained by the HRF alone. The functions were convolved with the stimulus onsets to create covariates in a general linear model. Parameter estimates for the HRF regressor were calculated from the least mean squares fit of the model to the time series. Parameter estimates for the temporal derivative were not considered in any contrast. Incorrect responses were calculated as a parameter estimate of no interest. For every subject the contrasts testblr>control and testscr>control were calculated. In a random effects group analysis these contrasts were subjected to a paired t-test between the variables scrambled (scr) and blurred (blr). Voxels with a significance level of $p < 0.001$ uncorrected belonging to clusters with at least 5 voxels are reported.

6.2.3 Results

Behavioural data: The mean accuracy rate was 70.54% in the blurred condition, 80.58% in the scrambled condition, and 90.18% in the control condition. Pair-wise comparisons revealed that in the scrambled condition participants performed slightly better than in the blurred condition, $p = .021$. The accuracy in the control condition was significantly higher than the blurred condition, $p < .001$, and also higher than the scrambled condition, $p < .05$. The mean RT for blurred trials was 1133 ms (*STD* = 122.88), for scrambled trials the mean RT was 1184 ms (*STD* = 134.44), in the control condition the mean RT was 916 ms (*STD* = 93.86). Pair-wise comparisons showed no difference between scrambled and blurred condition ($p = .559$), but the RTs of the control condition were marginally shorter in the control condition that in the experimental conditions (blurred: $p = .058$, scrambled: $p = .073$).

Table 1. Peak activations in blurred to scrambled contrast. Cerebral areas with corresponding Brodman areas (BA), Z-values and mni-coordinates for these peaks are reported.

Cerebral area	side	BA	z-value	coordinates (mni)	cluster size
Superior frontal gyrus	R	10	4.58	24, 57, 27	6
		6	3.96	9, 15, 63	29
	L	6	3.50	-6, 6, 66	
		6	3.82	-9, 21, 60	17
Medial frontal gyrus	L	32	3.72	-21, 48, 15	7
		32	3.57	-6, 18, 48	9
Inferior frontal gyrus	R	47	3.44	48, 33, -3	12
Middle temporal gyrus	L	39	3.53	-51, -57, 3	6
	R	21	3.39	60, -54, 0	9

fMRI Data: The paired t-test of testblr and testscr faces elicited a significant BOLD fMRI signal difference. Blurred trials as opposed to scrambled trials revealed activation in the middle temporal gyrus bilaterally.

Additionally, the right and left superior frontal gyrus and the left medial frontal gyrus were activated by configural trials (see Table 1). Scrambled trials activated the left posterior fusiform gyrus, left precuneus, areas of the left parietal lobe, the left lingual gyrus and the right insula (see Table 2). Figure 15 illustrates the differential activation between blurred and scrambled processing.

6.3 Discussion

In the present study brain regions were traced that were selectively active for configural (blurred trials) and featural (scrambled trials) face processing. I found bilateral activation of the middle temporal gyrus during configural face processing. Featural processing selectively activated the left fusiform gyrus, parietal lobe, lingual gyrus, and precuneus. Furthermore, the right insula was activated during featural face processing. Because the visual information contained in the critical stimuli was the same for configural and featural processing, this difference of activation can only be due to a different mechanism used for processing featural and configural information. The differential neural activation of featural and configural processing found in the present study is compatible with the dual-code view of face processing often suggested in previous studies (chapter 2; see also Bartlett et al., 2003; Cabeza & Kato, 2000; Schwaninger et al., 2002).

Table 2. Peak activations in scrambled to blurred contrast. Cerebral areas with corresponding Brodman areas (BA), Z-values and mni-coordinates for these peaks are reported.

Cerebral area	side	BA	z-value	coordinates (mni)	cluster size
Precuneus	L	7	3.91	-15, -63, 36	8
Insula	R	13	3.79	36, -42, 21	5
Fusiform gyrus	L	37	3.75	-33, -54, -12	10
Parietal lobe	L	31	3.74	-21, -51, 36	10
Lingual gyrus	L	19	3.64	-27, -69, -3	6

Left hemisphere activation when applying a featural processing mode is in line with findings of Martinez and colleagues (Martinez et al., 1997), who found hemispheric differences for local and global processing (see also Rossion et al., 2000).

Andrews and Ewbank (2004) found evidence that face selective regions within the inferior temporal lobe are involved in the perception and recognition of faces, while processing of changeable aspects of faces (e.g., different viewpoints of faces and facial expression) is associated with superior temporal face-selective regions. The present data suggest that configural and feature based processes embrace different regions than the inferior temporal lobe and in particular the FFA. The regions found for configural and featural processing are situated posterior to the fusiform face area. These findings are in line with the idea that featural and configural information is extracted from the input representation of a face in the primary visual cortex in the bottom-up course of the visual stream. These separate pathways then project to the FFA, were featural and configural information is combined to "holistic" face representations (c.f. Schwaninger et al.,

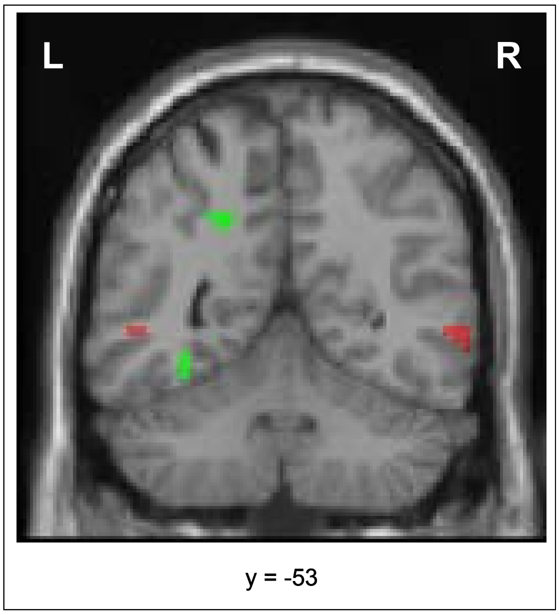

Figure 15: Differential processing of featural (green) and configural (red) face information. The activation map (p<.001, uncorrected, minimal cluster size 5 voxels) is shown superimposed onto a selected coronal slice of the EPI-template provided by SPM2. The section was taken coronally; the anterior-posterior level is based on mni coordinates.

2002).

Insofar the fact that the FFA showed no activation here may seem surprising only at first view. Given that the critical stimuli were intact faces in both conditions, it is plausible that regions selective for faces per se were subtracted in the contrasts at hand. Only activation that was selective for featural or configural processing remained. Activation of the FFA was typically found when faces were contrasted to objects or deranged faces. Here, intact faces were contrasted to intact faces, the only difference being the information given prior to the test face. The task could only be solved by using either configural or featural information to process the same intact face. Thus, the activation revealed by the present contrasts constitutes the processing mode adopted to encode the test face, suggesting that FFA is not specifically involved in configural and featural processing. This assumption is not consistent with the findings of Rossion and colleagues (Rossion et al., 2000), who reported a double dissociation between configural and featural processing modes within the FFA. This discrepancy may be the consequence of different paradigms. While in the study of Rossion et al. (2000) participants had to attend to either eyes, mouth, or the whole face and indicate whether the parts or the whole faces were the same. Possibly, the findings of Rossion et al. (2000) reflect different attentional strategies instead of configural and featural processing. The participants here had to match and therefore recognize a face on the basis of configural or featural information they just saw before. A further difference between these two studies is that Rossion et al. (2000) used a block-design in a PET-study, whereas here I used fMRI using an event-related design. Finally, Rossion et al. (2000) analysed the percentage of blood flow changes only within the right and left FFA. I was interested in the whole brain activity of featural and configural processing and did not restrict my analyses to face specific regions.

The behavioural data suggest that configural and featural tasks were not of equivalent difficulty, as scrambled trials were solved more accurately than blurred trials. This imbalance in task difficulty reflects findings of a previous behavioural study using a similar design, where scrambled faces were matched more easily than blurred faces (Experiment 2). Most other studies reported preponderance for configural face information (e.g., Cabeza & Kato, 2000; Schwaninger et al., 2002). These studies however used learnt faces as stimuli, suggesting that configural information seems to be crucial for recognition of familiar faces. In contrast, featural information seems to be particularly important for recognizing novel faces (chapter 2). Because the faces in the present study were not learnt beforehand, my results reflect novel face perception. It could be argued that the differences in our BOLD responses could be related to differences in task difficulty. If so, more activated voxels in the visual areas would be expected, as more brain structures would be expected to be involved. However, the opposite was the case. Only 15 voxels were activated in the blurred trials compared to more than 26 in the scrambled trials. This indicates that the BOLD responses measured with fMRI can not be directly associated with behavioural measures, such as reaction times and accuracy (cf. McGonigle et al. 2000). However, the frontal regions that were activated during blurred trials reflect this unbalanced task difficulty.

The data presented here therefore clearly suggest a dual-code view where featural and configural information is processed following separate pathways. Whether these pathways coincide with the ventral stream ("what-system") and dorsal stream ("where system") (Haxby et al., 1991; Ungerleider & Mishkin, 1982) or with hemispheric differences is not clearly apparent from the present data. The data suggest that featural processing is indeed lateralized whereas configural processing occurs bilaterally. Similarly, some of the regions processing featural information are indeed located ventral to the middle temporal gyrus, which showed more activation for configural processes. But at the same time featural processing activated a region parietal lobe, which lies dorsal to the middle temporal gyrus. Whether the dissociation found here is face specific or specific to featural and configural processes in general will have to be the issue of future studies. Against the background of distributed neural systems for domain specific processing it is conceivable that this dissociation of configural and featural processes will be found in all kind of visual recognition.

7. Synopsis: Where does all this leave us?

Abstract

In the following chapter the most important findings of chapters 2 – 6 are summarised and integrated into a new model of face processing. Furthermore, theoretical issues from chapter 1 are picked up and discussed anew. On the basis of the present findings the holistic hypothesis of face processing is defeated and instead, the featural-configural hypothesis relishes updraught. Finally, a few limitations of the present studies are discussed and issues for further research are suggested.

7.1 Introduction

The experiments reported in Chapters 2-6 help to understand the mechanisms involved in face recognition. The role of isolated featural and configural information was investigated by using scrambled and blurred faces as stimuli. Featural (or component, piecemeal) information is understood as the information contained in the facial parts, such as the eyes, nose and mouth. Configural information denotes the information contained in the spatial inter-relationship of the parts (see chapter 1 for further information on featural and configural processing). By scrambling a face into its constituent parts the detail information contained in the parts is preserved, while the spatial inter-relationship between them is disrupted. By blurring the faces with a low-pass filter the particular detail information is disrupted while the spatial inter-relationship between the parts is conserved. With this method featural and configural processing strategies could be activated separately and conclusions could be made about the nature of configural and featural representations. Overall, the studies challenge the view that faces are perceived, processed and stored as wholes.

7.2 The defeat of the holistic hypothesis

A consistent outcome of all the experiments presented here is that a pure holistic view where faces are processed and stored as unparsed wholes must be rejected. Specifically the experiments presented in chapters 2, 3, and 6 show that faces are not processed holistically, not exclusively anyway. Holistic face representations may well be found in the fusiform face area, but along the visual pathway featural and configural information seems to be processed following separate routes. In the experiments reported in chapter 2 scrambled and blurred faces could be reliably matched to intact whole faces. This finding gives evidence that both isolated featural and isolated configural information could be processed. A pure holistic view of face processing would not allow that faces can be recognized on the basis of the parts. However, in all experiments presented here featural information sufficed to recognize a face. In fact, in the experiments using novel faces featural information even seems to play a more important role than configural information. The often reported predominance of configural information seems to hold only for faces which have been encoded longer than a second. In the first instance of seeing a novel face featural processing strategies seem to preponderate. As soon as a face is sufficiently studied, even during a very short time of a couple of seconds, configural representations can be built of the face and the importance of configural information increases. This could be demonstrated in Experiment 3, where half of the faces were intensively studied before the Experiment resulting in relatively better recognition of blurred faces. The featural-configural hypothesis of face perception found further support in the fact that no inversion effect was found for scrambled faces. Blurred faces however evoked a strong inversion effect. This differential sensitivity to inversion substantiates the presumption that configural and featural processing constitute two separate routes to face processing and is in line with findings of previous studies (e.g., Leder and Bruce, 2000; Sergent, 1984), which showed that configural changes were affected by inversion more that featural changes. The comprehension that faces are processed along two separate pathways is of computational interest when it comes to developing and enhancing electronic face recognition devices. All algorithms attempting to obtain similar performance in face recognition as normal human beings will coercively have to include separate processes for featural and configural information.

The distinction of featural and configural processing modes has lead to the hypothesis claiming that faces are special in the way that they are processed predominantly on the basis of configural information as opposed to other objects which are reported to be processed part-based. In accordance with this hypothesis, the results of Experiment 6 revealed that houses were less sensitive to inversion than faces. Regardless of the fact that houses are objects and thus are processed more part-based, this finding is noteworthy because houses, just like faces, are made up of a number of different parts (windows, doors, etc.) which are organized spatially to form an individual configuration. Moreover, Thatcherized houses, where the windows and doors were inverted within the whole house to create stimuli similar to thatcherized faces, failed to show a Thatcher effect comparable to faces. As the Thatcher effect is the result of the interrelation of featural and configural information on the one hand and inversion on the other, this finding suggests that configural information is not so important for house processing. While this finding is of special interest for research on object recognition, it suggests that faces are unique in the way that a) configural information is indeed more important for face than for object recognition and b) that the Thatcher illusion seems to be restricted to faces.

7.3 Towards a new model of face perception

Bruce and Young (1986) presented an influential model of face perception. In their model visual input undergoes structural encoding processes providing more abstract visual descriptions. These descriptions then project to the so-called face recognition unit (FRU). The FRU sends signals of resemblance to decision processes within the cognitive system. The cognitive system and the FRU then feed to the person identification nodes (PIN).

The findings reported in the present thesis enhance the model proposed by Bruce and Young (1986) and also specify the model recently proposed by Schwaninger et al. (2002). A

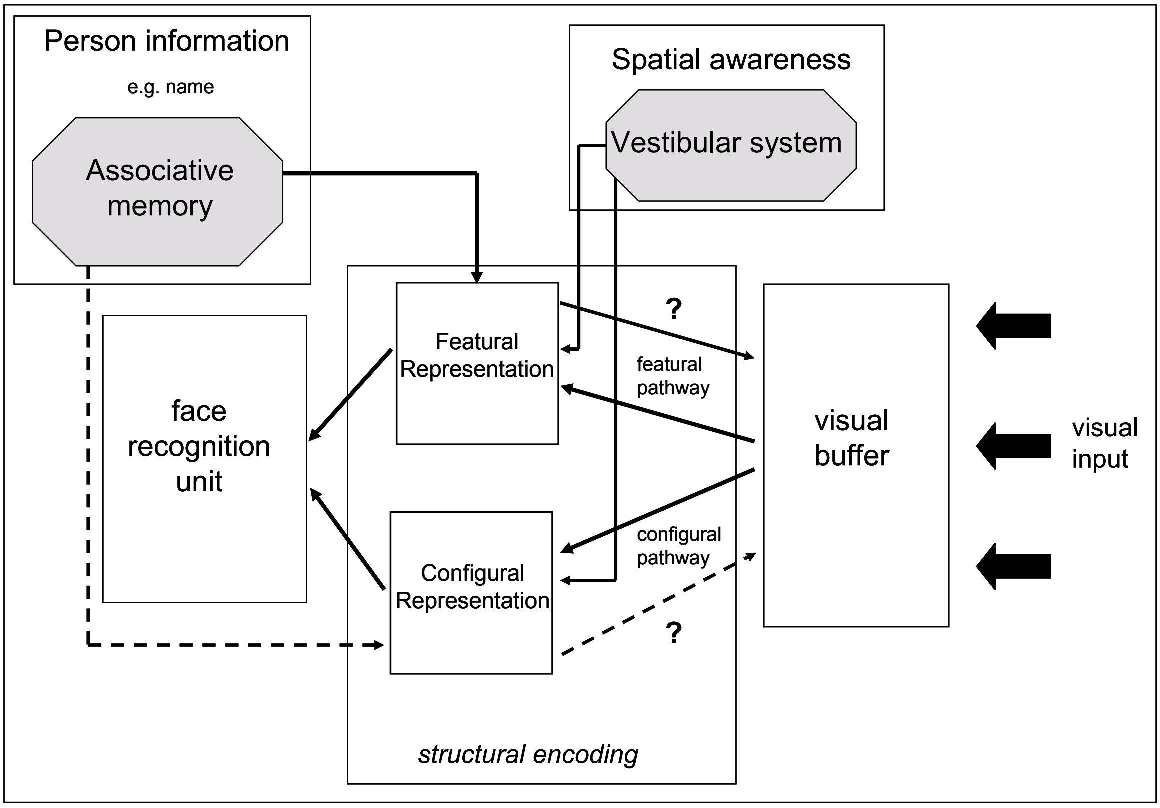

Figure 16: New integrative model for face recognition: Visual input is first represented in the visual buffer (presumably located in primary visual cortex), from which featural and configural information is extracted and processed following separate pathways. In the face recognition unit these featural and configural representations are combined into a ("holistic") representation of the whole face. During structural encoding the vestibular system feeds into the configural and featural representations providing information about the spatial orientation of the face with respect to gravity. Top-down activation of a face via visual mental imagery activates both configural and featural representations, albeit featural representations more than configural representations.

new model of face recognition where the present findings are integrated is shown in Figure 16. Specifically, the structural encoding processes can be characterized more precisely: The abstract visual description proposed by Bruce and Young (1986) can be differentiated in featural and configural processing. These processing types follow two separate pathways, as was substantiated in behavioural (Experiments 1-3) and neuro-imaging (Experiment 7) studies. Faces could be reliably recognised on the basis of isolated configural and featural information. In the model proposed here it is assumed that featural and configural information is extracted from the input representation in the visual buffer (presumably located in primary visual cortex), to form featural and configural representations. Brain structures involved in configural processing include the middle temporal gyrus. Featural processing takes place in the left posterior fusiform gyrus, left precuneus, areas of the left parietal lobe, and the left lingual gyrus. The featural and configural representations then feed forward to the face recognition unit (FRU), where they are combined into representations of the whole face. The FRU is presumably located in the FFA and the representations in the FRU may have holistic characteristics. This would explain why most neuro-imaging studies on face processing found activation in the FFA whereas Experiment 7 did not. In Experiment 7 whole faces where contrasted with whole faces, the only difference was that in one case a featural processing mode was activated and a configural processing mode was activated in the other. The activation found here revealed regions that are responsible for configural processing or featural processing, respectively. Therefore it is not unexpected that I did not find activation in areas dedicated to whole face processing.

Experiments 1-3 further revealed that in novel face recognition featural information preponderates, but during the process of familiarization configural information becomes relatively more important. Further, as the results of Experiment 4 showed, featural and configural representations can be activated bottom-up via visual perception and top-down via visual mental imagery. Top-down activated featural and configural representations then activate the face recognition units. However, visual mental imagery of learnt faces seems to activate primarily featural representations, whereas perception of the same faces activates featural and configural representations to similar extent. Whether activation of configural and featural representations feed back to primary visual cortex as hypothesized in the model proposed here and as would be assumed by Kosslyn will have to be explored in future studies.

In connection with the face inversion effect Experiment 5 yielded interesting findings. It seems that configural and featural representations are not entirely independent of gravitational direction. Specifically, thatcherized faces were more difficult to detect when the observer was tilted 135° along the body roll axis. This suggests that not only the retinal orientation of the stimulus itself affects the FIE, but also the general orientation within the gravitational frame of reference. As for the proposed model this suggests that apart from input from primary visual cortex and top-down activation via mental imagery, configural and featural representations are also fed with information from regions computing gravitational spatial orientation. Since the Thatcher illusion specifically involves the interrelation of featural and configural information and Experiment 5 showed an interaction of body tilt and stimulus orientation, it seems plausible to assume a connection between the vestibular system and featural and configural representations. Whether the vestibular system also feeds into the primary visual representation or into the face recognition unit goes beyond the limits of the present findings.

7.4 Limitations

Face processing usually involves face identification. When we see a face we usually immediately know that it is Tom, or the postman, or the lady-who-lives-at-number-four. Similarly, when we look through glossy magazines we recognize photographs as Angelina Jolie or David Beckham. The act of face recognition therefore requires that we somehow get from a visual pattern (a face) to a semantic level of representation where the identity of the face is specified (cf. Bruce, 1988). In real life situations familiar faces are coded together with facts about the people such as their occupation, their hobbies, their marital status, age, address, etc. In contrast, face identification in experimental conditions mostly involves correctly recalling faces that have been studied beforehand. To which extent semantic information is used to identify faces has not been addressed in the studies presented in this thesis. When familiar faces were an issue (Experiments 3 and 4), a set of faces were studied beforehand, sometimes together with a name (Experiment 4). Personal information about the studied faces was not provided. Although this might restrict conclusions about familiar face identification to a certain extent, the studies presented here manage to shed light on the perceptual aspects of face identification.

Like most studies on face perception, the experiments reported here were conducted with front-view, two-dimensional, still pictures of neutral faces. However, faces are a complex three-dimensional class of objects and have a varying appearance depending on the view of the face. This is a problem which has been often neglected. Even Penry, the inventor of Photofit, described how faces appear balanced or unbalanced in terms of their proportion, but these proportions are derived entirely through consideration of faces as two-dimensional patterns (cf. Bruce, 1988). As to the use of front-view faces nothing can be said about view-specific processing of faces on the basis of the present studies. All claims made here hold true for front view faces, whether the present findings can be generalized to other views will have to be the issue of future studies.

The different parts of a face most likely do not share the same salience. For example, information contained in the eyes may be more concise than information contained in the nose. This may result in observers paying more attention to the eyes than to the nose in the scrambled conditions. Also, the conciseness of the parts may vary between individual faces. This was not controlled for in the present experiments. However, as all features were presented in the scrambled stimuli, differential importance of the individual features was not an issue. The participants could choose the most significant feature to solve the task.

The present thesis concentrates on featural and configural processes of face perception. The same distinction is conceivable for object recognition. However, object recognition has barely been considered here. Solely the results of Experiment 6 using thatcherized houses as stimuli indicate that the interrelationship of configural and featural information is not so cogent in houses. Yet, it would be interesting to further investigate configural and featural processing in objects, especially against the background of the specifity of face processing. Planned studies will investigate scrambled and blurred common objects and will complement the findings of the present studies.

7.5 Conclusion

The studies presented in this thesis revealed that in face processing both configural and featural information play an important role and the processing of both information types follows two distinct pathways. Specifically, Experiments 1-3 and 7 give evidence that faces can be recognized when only one pathway is activated. That is, faces could be recognized when only configural or only featural information was provided. On the basis of these findings the model of face perception presented by Bruce and Young (1986) was specified and extended. From the input representation in primary visual cortex featural and configural information is extracted generating featural and configural representations. These representations can be activated bottom-up (Experiments 1-4 and 7) and top-down (Experiment 4). Furthermore, information from the vestibular system is fed into these representations (Experiment 5). Experiment 6 suggests that at least some aspects of face processing are face-specific, such as the Thatcher illusion. These findings are of computational interest when it comes to developing and enhancing electronic face recognition devices.

8. Glossary

Blurred faces: Manipulation employed to separately investigate ➔ *configural information*. A low-pass filter is applied, resulting in the stimuli appearing blurred.

d' (D-prime): Measure of signal detectability, equal to the difference between the z-transformed hit rates and false alarm rates. A *d'* significantly larger than 0 means that a signal could be detected above chance level.

Configural information: Information contained in the spatial inter-relationship of the features.

Dual mode view: View that both featural information and configural information play an important role in face processing and that featural and configural processing follow two distinct pathways.

Face inversion effect (FIE): Phenomenon that faces appear to be much more sensitive to inversion than other object classes.

Face recognition unit (FRU): assumptive theoretical location in the brain where face recognition occurs (cf. Bruce & Young, 1986; Bruce, 1988)

Featural information: Information contained in the constituent parts of a face. Featural information is used interchangeably with piecemeal information or component information.

First-order relational information: (= Categorical spatial relations). Basic arrangement of the facial parts (e.g., the nose lies between the eyes, above the mouth; cf. Diamond & Carey, 1986).

Frame of reference: Set of standards with respect to which the properties of visual objects can be described.

Fusiform face area (FFA): Area in the fusiform gyrus which shows activation when faces are presented in brain-imaging experiments such as fMRI. Because the FFA is particularly responsive to faces it has been assumed that the FFA is a face specific region.

Gestalt: Visual apprehension where the perception of the whole takes precedence over the sum of its constituent parts, in the sense of "the whole is different than the sum of its parts".

Greebles: Computer-generated novel class of objects which are comparable to faces, as all exemplars share the same number of parts in the same first-order configuration.

Holistic face processing: View that faces are processed and stored as unparsed wholes, without explicitly representing the facial parts, comparable to a bitmap.

IdentiKit faces: Dunleavy (1959; 1975). Facial composite construction system used by police forces to find criminals. The use of this system involves the selection of individual photographic parts from a kit, and then these component features are assembled into a whole face configuration. American version of the British ➔ *Photofit* system.

Mental rotation: Process by which people are able to continuously change the orientation of an object in imagination.

Oblique effect: Observation that performance of a visual task is superior when the stimuli are oriented vertically or horizontally compared to when they are obliquely oriented.

Other race effect: Observation that faces of a foreign racial group seem to resemble each other more than faces of our own racial group.

Person identification node (PIN): Assumptive theoretical location in the brain where person recognition (as opposed to face recognition) occurs (cf. Bruce & Young, 1986; Bruce, 1988)

Photofit faces: Penry (1971). Facial composite construction system used by police forces to find criminals. The use of this system involves the selection of individual photographic parts from a kit, and then these component features are assembled into a whole face configuration. British version of the US American ➔ *IdentiKit* system

Prosopagnosia: "Face-blindness". People suffering from prosopagnosia are incabable of recognizing people by their faces. Often prosopagnosia arises after lesions in the FFA, but there is growing evidence of a congenital form of prosopagnosia.

Schema: Organizing mechanism for information input and output which develops and improves according to experiences gained through interaction with incoming stimuli.

Scrambled faces: Manipulation employed to separately investigate ➔ *featural information*. The facial parts are cut out and presented in a random arrangement, resulting in a loss of configural information while maintaining featural information.

Second-order relational information: (= Metric spatial relations). Metric distances between the facial parts (cf. Diamond & Carey, 1986).

Thatcher illusion: Illusion first described by Thompson (1980) where a face appears highly grotesque when eyes and mouth are inverted within the whole upright face. When viewed upside-down, such a face looks rather normal.

Visual buffer: Spatially organized array within the visual system that constitutes the medium for visual images (cf. Kosslyn, 1994). It is suggested to be located within primary visual cortex.

9. References

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage, 23(3)*, 905-913.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104*, 54 - 75.

Bartlett, J. C., Searcy, J., & Abdi, H. (2003). What are the routes to face recognition? In M. A. Petersen & G. Rhodes (Eds.), *Perception of faces, objects, and scenes* (pp. 21-52). New York: Oxford University Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94(2)*, 115-147.

Bodamer, J. (1947). Die Prosop-Agnosie. *Arch. Psychiat. Nervenkrankheit, 179*, 6-53.

Bruce, V. (1988). *Recognizing faces*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology, Applied, 7*, 207 - 218.

Buchanan-Smith, H. M., & Heeley, D. W. (1993). Anisotropic axes in orientation perception are not retinotopically mapped. *Perception, 22(12)*, 1389-1402.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10(3)*, 243-248.

Buttle, H., & Raymond, J. E. (2003). High familiarity enhances visual change detection for face stimuli. *Percept Psychophys, 65(8)*, 1296-1306.

Cabeza, R., Burton, A. M., Kelly, S. W., & Akamatsu, S. (1997). Investigating the relation between imagery and perception: evidence from face priming. *Q J Exp Psychol A, 50(2)*, 274-289.

Cabeza, R., & Kato, T. (2000). Features are also important: contributions of featural and configural processing to face recognition. *Psychol Sci, 11(5)*, 429-433.

Charcot, J.-M., & Bernard, D. (1883). Un cas de suppression brusque et isolée de la vision mentale des signes et des objets (formes et couleurs). *Le Progrès Médical, 11*, 568-571.

Clément, G., & Eckhardt, J. (2005). Influence of the gravitational vertical on geometric visual illusions. *Acta Astronautica, 56*, 911-917.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257-271.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception, 29(8)*, 893-909.

Collishaw, S. M., & Hole, G. J. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception, 31(3)*, 287-296.

Corballis, M. C., Nagourney, B. A., Shetzer, L. I., & Stefanatos, G. (1978). Mental rotation under head tilt: factors influencing the location of the subjective reference frame. *Percept Psychophys, 24(3)*, 263-273.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience, 3*, 1-8.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. J *Exp Psychol Gen, 115(2)*, 107-117.

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory and Cognition, 25(2)*, 129-139.

Farah, M. J., Peronnet, F., Gonon, M. A., & Giard, M. H. (1988). Electrophysiological evidence for a shared representational medium for visual images and visual percepts. J *Exp Psychol Gen, 117(3)*, 248-257.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? J *Exp Psychol Hum Percept Perform, 21(3)*, 628-634.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychol Rev, 105(3)*, 482-498.

Freire, A., & Lee, K. (2001). Face recognition in 4- to 7-year-olds: Processing of configural, featural, and paraphernalia information. *Journal of Experimental Child Psychology, 80(4)*, 347-371.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *Neuroimage, 7(1)*, 30-40.

Gaunet, F., & Berthoz, A. (2000). Mental rotation for spatial environment recognition. *Brain Res Cogn Brain Res, 9(1)*, 91-102.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci, 3(2)*, 191-197.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Res, 37(12)*, 1673-1682.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci, 2(6)*, 568-573.

Goldstein, A. G. (1975). Recognition of Inverted Photographs of Faces by Children and Adults. *Journal of Genetic Psychology, 127(1)*, 109-123.

Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. *The Journal of Psychology, 105*, 47-59.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci, 7(5)*, 555-562.

Grosbras, M. H., Leonards, U., Lobel, E., Poline, J. B., LeBihan, D., & Berthoz, A. (2001). Human cortical networks for new and familiar sequences of saccades. *Cereb Cortex, 11(10)*, 936-945.

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception, 13(5)*, 505-512.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293(5539)*, 2425-2430.

Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., et al. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc Natl Acad Sci U S A, 88(5)*, 1621-1625.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn Sci, 4(6)*, 223-233.

Henderson, Z., Bruce, V., & Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology, 15(4)*, 445-464.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci, 3(1)*, 80-84.

Hole, G. J. (1994). Configurational Factors in the Perception of Unfamiliar Faces. *Perception, 23(1)*, 65-74.

Ishai, A., Haxby, J. V., & Ungerleider, L. G. (2002). Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *Neuroimage, 17(4)*, 1729-1741.

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed Neural Systems fort he Generation of Visual Images. *Neuron, 28*, 979-990.

Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *J Cogn Neurosci, 12 Suppl 2*, 35-51.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A, 96(16)*, 9379-9384.

Jaggi-Schwarz, K., & Hess, B. J. (2003). Common reference system for estimation of the postural and subjective visual vertical. *Ann N Y Acad Sci, 1004*, 516-520.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40(1-2)*, 1-19.

Joseph, J. E., & Gathers, A. D. (2002). Natural and manufactured objects activate the fusiform face area. *Neuroreport, 13(7)*, 935-938.

Kanamori, N., & Yagi, A. (2002). The difference between flipping strategy and spinning strategy in mental rotation. *Perception, 31(12)*, 1459-1466.

Kanwisher, N. (2000). Domain specificity in face perception. *Nat Neurosci, 3(8)*, 759-763.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci, 17(11)*, 4302-4311.

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport, 10(1)*, 183-187.

Kaptein, R. G., & Van Gisbergen, J. A. (2004). Interpretation of a discontinuity in the sense of verticality at large body tilt. *J Neurophysiol, 91(5)*, 2205-2214.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, Massachusetts: The MIT Press.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 51(3)*, 449-473.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 53(2)*, 513-536.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception, 30(1)*, 73-83.

Leube, D. T., Yoon, H. W., Rapp, A., Erb, M., Grodd, W., Bartels, M., et al. (2003). Brain regions sensitive to the face inversion effect: a functional magnetic resonance imaging study in humans. *Neurosci Lett, 342(3)*, 143-146.

Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. J *Neurosci, 20(2)*, 878-886.

Levine, D. N., Warach, J., & Farah, M. (1985). Two visual systems in mental imagery: dissociation of "what" and "where" in imagery disorders due to bilateral posterior cerebral lesions. *Neurology, 35(7)*, 1010-1018.

Lewis, M. B. (2001). The lady's not for turning: rotation of the Thatcher illusion. *Perception, 30(6)*, 769-774.

Lipshits, M., Bengoetxea, A., Cheron, G., & McIntyre, J. (2005). Two reference frames for visual perception in two gravity conditions. *Perception, 34(5)*, 545-555.

Lipshits, M., & McIntyre, J. (1999). Gravity affects the preferred vertical and horizontal in visual perception of orientation. *Neuroreport, 10(5)*, 1085-1089.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends Cogn Sci, 4(1)*, 6-14.

Lobmaier, J. S., & Mast, F. W. (in press). The Thatcher illusion: Rotating the viewer instead of the picture. *Perception*.

Macho, S., & Leder, H. (1998). Your eyes only? A test of interactive influence in the processing of facial features. *J Exp Psychol Hum Percept Perform, 24(5)*, 1486-1500.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing Orientation and Face Recognition. *Psychological Science, 13(2)*, 194-196.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17-24.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Martinez, A., Moses, P., Frank, L., Buxton, R., Wong, E., & Stiles, J. (1997). Hemispheric asymmetries in global and local processing: evidence from fMRI. *Neuroreport, 8(7)*, 1685-1689.

Mast, F. W. (2000). Does the world rock when the eyes roll? Allocentric orientation representation, ocular counterroll, and the subjective visual vertical. *Swiss Journal of Psychology, 59*, 89-101.

Mast, F. W., Ganis, G., Christie, S., & Kosslyn, S. M. (2003). Four types of visual mental imagery processing in upright and tilted observers. *Cognitive Brain Research, 17*, 238-247.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6(6)*, 225-260.

Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cereb Cortex, 14(11)*, 1256-1265.

Michelon, P., & Zacks, J. M. (2003). What is primed in priming from imagery? *Psychol Res, 67(2)*, 71-79.

Mittelstaedt, H. (1999). The role of the otoliths in perception of the vertical and in path integration. *Ann N Y Acad Sci, 871*, 334-344.

Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing skills. *J Exp Child Psychol, 86(1)*, 67-84.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31(5)*, 553-566.

Murray, J. E. (1997). Flipping and spinning: spatial transformation procedures in the identification of rotated natural objects. *Mem Cognit, 25(1)*, 96-105.

Murray, J. E. (2004). The ups and downs of face perception: Evidence for holistic encoding of upright and inverted faces. *Perception, 33(4)*, 387 - 398.

Nachson, I., & Shechory, M. (2002). Effect of inversion on the recognition of external and internal facial features. *Acta Psychol (Amst), 109(3)*, 227-238.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *J Cogn Neurosci, 12(6)*, 1013-1023.

Palermo, R., & Rhodes, G. (2002). The influence of divided attention on holistic face perception. *Cognition, 82(3)*, 225-257.

Pallis, C. A. (1955). Impaired identification of faces and places with agnosia for colours: Report of a case due to cerebral embolism. *J Neurol Neurosurg Psychiatry, 18*, 218-224.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos Trans R Soc Lond B Biol Sci, 335(1273)*, 23-30.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., et al. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Exp Brain Res, 86(1)*, 159-173.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res, 47(3)*, 329-342.

Prinzmetal, W., & Beck, D. M. (2001). The tilt-constancy theory of visual illusions. *J Exp Psychol Hum Percept Perform, 27(1)*, 206-217.

Rakover, S. S. (1999). Thompson's Margaret Thatcher illusion: when inversion fails. *Perception, 28(10)*, 1227-1230.

Rakover, S. S. (2002). Featural vs. configurational information in faces: a conceptual and empirical analysis. *Br J Psychol, 93(Pt 1)*, 1-30.

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition, 47(1)*, 25-57.

Rhodes, G., Byatt, G., Michie, P. T., & Puce, A. (2004). Is the fusiform face area specialized for faces, individuation, or expert individuation? *J Cogn Neurosci, 16(2)*, 189-203.

Rhodes, G., Tan, S., Brake, S., & Taylor, K. (1989). Expertise and configural coding in face recognition. *Br J Psychol, 80 (Pt 3)*, 313-331.

Robbins, R., & McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition, 88(1)*, 79-107.

Robertson, L. C., Lamb, M. R., & Knight, R. T. (1988). Effects of lesions of temporal-parietal junction on perceptual and attentional processing in humans. *J Neurosci, 8(10)*, 3757-3769.

Rock, I. (1973). *Orientation and Form*. New York: Academic Press.

Rock, I. (1988). On Thompson's inverted-face phenomenon (Research Note). *Perception, 17*, 815 - 817.

Rossion, B., Dricot, L., Devolder, A., Bodart, J. M., Crommelinck, M., De Gelder, B., et al. (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *J Cogn Neurosci, 12(5)*, 793-802.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Rewiews, 1(1)*, 62-74.

Rossion, B., Schiltz, C., & Crommelinck, M. (2003). The functionally defined right occipital and fusiform "face areas" discriminate novel from visually familiar faces. *Neuroimage, 19(3)*, 877-883.

Rossion, B., Schiltz, C., Robaye, L., Pirenne, D., & Crommelinck, M. (2001). How does the brain discriminate familiar and unfamiliar faces?: a PET study of face categorical perception. *J Cogn Neurosci, 13(7)*, 1019-1034.

Sagiv, N., & Bentin, S. (2001). Structural encoding of human and schematic faces: holistic and part-based processes. *J Cogn Neurosci, 13(7)*, 937-951.

Schöne, H. (1964). On the role of gravity in human spatial orientation. *Aerospace Medicine, 35*, 764-772.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36-71.

Schwaninger, A., Carbon, C.-C., & Leder, H. (2003). Expert face processing: Specialization and constraints. In G. Schwarzer & H. Leder (Eds.), *Development of face processing* (pp. 81-97). Göttingen: Hogrefe.

Schwaninger, A., Lobmaier, J. S., & Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. *Lecture Notes in Computer Sciences, 2525*, 634-650.

Schwaninger, A., Lobmaier, J. S., & Fischer, M. H. (2005). The inversion effect on gaze perception reflects processing of component information. *Exp Brain Res, 167(1)*, 49-55.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relation information of faces. *Journal of Experimental Psychology: Human Perception and Performance, 22(4)*, 904-915.

Sergent, J. (1984). Configural processing of faces in the left and the right cerebral hemispheres. *J Exp Psychol Hum Percept Perform, 10(4)*, 554-572.

Simons, D. J., Wang, R. F., & Roddenberry, D. (2002). Object recognition is mediated by extraretinal information. *Percept Psychophys, 64(4)*, 521-530.

Stevenage, S. V. (1995). Can Caricatures Really Produce Distinctiveness Effects. *British Journal of Psychology, 86*, 127-146.

Sturzel, F., & Spillmann, L. (2000). Thatcher illusion: dependence on angle of rotation. *Perception, 29(8)*, 937-942.

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *J Exp Psychol Gen, 130(3)*, 534-543.

Tanaka, J. W., & Farah, M. (2003). The holistic representation of faces. In M. A. Petersen & G. Rhodes (Eds.), *Perception of faces, objects, and scenes* (pp. 53-74). Oxford: Oxford University Press.

Tanaka, J. W., & Farah, M. J. (1991). Second-order relational properties and the inversion effect: testing a theory of face perception. *Percept Psychophys, 50(4)*, 367-372.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Q J Exp Psychol A, 46(2)*, 225-245.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Mem Cognit, 25(5)*, 583-592.

Tanaka, J. W., & Taylor, M. (1991). Object Categories and Expertise - Is the Basic Level in the Eye of the Beholder. *Cognitive Psychology, 23(3)*, 457-482.

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception, 9(4)*, 483-484.

Troje, N. F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception, 32(2)*, 201-210.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113*, 169-193.

Udo de Haes, H. A. (1970). Stability of apparent vertical and ocular countertorsion as a function of lateral tilt. *Perception & Psychophysics, 8*, 137-142.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. G. Ingle, M. A. Goodale & R. J. Q. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.

Valentine, T. (1988). Upside-down faces: A review of the effects of inversion upon face recognition. *British Journal of Psychology, 79*, 471-491.

Valentine, T., & Bruce, V. (1985). What's up? The Margaret Thatcher illusion revisited. *Perception, 14(4)*, 515-516.

Van Beuzekom, A. D., & Van Gisbergen, J. A. M. (2000). Properties of the internal representation of gravity inferred from spatial-direction and body-tilt estimates. *Journal of Neurophysiology, 84*, 11-27.

Weissman, D. H., & Woldorff, M. G. (2005). Hemispheric Asymmetries for Different Components of Global/Local Attention Occur in Distinct Temporo-parietal Loci. *Cereb Cortex, 15(6)*, 870-876.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational Information in Face Perception. *Perception, 16(6)*, 747-759.

Young, A. W., Humphreys, G. W., Riddoch, M. J., Hellawell, D. J., & de Haan, E. H. (1994). Recognition impairments and face imagery. *Neuropsychologia, 32(6)*, 693-702.

Young, A. W., & Van De Wal, C. (1996). Charcot's case of impaired imagery. In C. Code, C.-W. Wallesch, A. R. Lecours & Y. Jeanette (Eds.), *Classic cases in Neuropsychology*. Hove: Lawrence Erlbaum.

Acknowledgements

There are many people I wish to thank for helping and supporting me in various phases of my PhD. First and foremost I gratefully acknowledge my debt to my supervisor and boss Fred Mast. During the three years of my PhD his advice was always competent and professional. At the same time he allowed me great topical freedom. Fred, it was a very fruitful and instructive time and I hope to continue collaborating with you. Thank you very much!

I am also grateful to my co-referee Bernhard Hess and to Alumit Ishai, who were both members of my steering committee within the PhD programme of the Center of Neuroscience Zurich (ZNZ).

A special thank you goes to my colleagues Peter Klaver and Tino Zähle who advised me on the analysis of the fMRI data. “Cheers!” also for the many stimulating coffee-breaks. Lastly, I would like to thank all the employees at Treichlerstrasse 10 for the pleasant atmosphere.

My research was supported by a grant awarded to Fred Mast from the Swiss National Science Foundation (Project No. 611-066052).

Curriculum Vitae

Personal Information

Date of Birth	2 March1974, in Port Elizabeth, South Africa
Hometown	Zurich
Marital status	single
Konfession	Roman catholic
Parents	Andreas Lobmaier (1949; electrician) Christine Lobmaier-Bednarska (1948; English teacher)

Current occupation:

January 2007 –	Postdoctoral fellow in St. Andrews, Scotland (Leader: Prof. Dave Perrett) (Grant from the Swiss National Science Foundation).
2003 – 2006	Research assistant and PhD Student at the Cognitive Neuroscience Lab, University of Zurich (Leader: Prof. Fred W. Mast)
Oktober 2006	Completion of PhD Thesis. *Faces in the mind: Featural and configural face representations in perception and imagery*.

Education:

2003 – 2006	International PhD programme in Neuroscience (Neuroscience Center, Zurich) http://www.neuroscience.unizh.ch/
1995 – 2003	Study of Psychology at the University of Zurich 1. Subsidiary Subject: Children and Youth Psychopathology 2. Subsidiary Subject: English linguistics Diploma thesis: *Featural and Configural Face Representations: Their Role in Familiar and Unfamiliar Face Recognition*.
2002	Internship at the University of Dundee, Scotland
2002	Practical training at the Kinder- und Jugendpsychiatrischer Dienst in Chur (Psychiatric centre for children and adolescents)
1994	Grammar school graduation (Stiftsschule Einsiedeln)